GENETIC ELEMENTS AND MOLECULAR MECHANISMS DRIVING THE EVOLUTION OF THE PATHOGENIC MARINE BACTERIUM *VIBRIO PARAHAEMOLYTICUS*

A Dissertation
Presented to
The Academic Faculty

by

Tracy Heather Hazen

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Biology

Georgia Institute of Technology
AUGUST 2009

GENETIC ELEMENTS AND MOLECULAR MECHANISMS DRIVING THE EVOLUTION OF THE PATHOGENIC MARINE BACTERIUM *VIBRIO PARAHAEMOLYTICUS*

Approved by:

Dr. Patricia A. Sobecky, Advisor
School of Biology
Georgia Institute of Technology

Dr. Roger Wartell
School of Biology
Georgia Institute of Technology

Dr. Thomas J. DiChristina
School of Biology
Georgia Institute of Technology

Dr. Eric V. Stabb
Department of Microbiology
University of Georgia

Dr. Jim Spain
School of Civil and Environmental Engineering
Georgia Institute of Technology

Date Approved: April 2009

ACKNOWLEDGEMENTS

I would like to thank my advisor Patricia Sobecky for her support. I am thankful to her for believing in me and supporting my interest in studying *Vibrios*. She agreed to my requests to apply for fellowships, attend conferences, and try new experiments. In addition, she encouraged me to become a part of the microbiology community nationally and internationally by sending me to numerous conferences, arranging for me to give presentations, and encouraging me to meet with the many speakers that she invited to give seminars at Georgia Tech.

There are numerous faculty and collaborators at Georgia Tech and other schools and research institutions that I owe thanks to. I would like to thank my thesis committee members Tom DiChristina, Jim Spain, Eric Stabb, and Roger Wartell for providing me with valuable feedback and guidance. Also, I would like to thank Rick Lovell and Julie Criminger at the University of South Carolina and Cheryl Bopp, Michele Parsons, Patti Lafon, and Nancy Garrett at the Centers for Disease Control and Prevention for their early collaborations that helped me on my way to studying *V. parahaemolyticus*. I would like to thank Soojin Yi at Georgia Tech for her collaboration and valuable instruction in molecular evolution. In addition, I would like to acknowledge the faculty of the Georgia Tech NSF IGERT program for their support by providing me with funding that allowed me to further my research and to present my findings at international conferences.

There are a number of fellow graduate students, scientists, postdocs, and staff that worked with me, provided technical guidance, and supported me with their friendship. I would like to thank Rob Martinez for all of his help throughout my time at Georgia Tech.

Also, Jason Dale, Karla Vincent, Jennifer Page, Kunmi Ayanbule, Seifu Seyoum, Yangfeng Chen, Justin Burns, and Frank Canella. I would especially like to thank Rupal Thazhath Cutting for being a great teaching mentor and supportive friend. In addition, I was fortunate to have worked with several talented and motivated undergrads, Danielle Kennedy, Daniel Silberger, Tiffany Lowe, and Shen Chen. They were a pleasure to work with and I know they will have great success.

Most importantly, I would like to thank my family. I thank my dad for sharing with me his love of microbiology, marine biology, and photography. His hard work and dedication to research have inspired me to set ambitious goals and to work to achieve them. I thank my mom who has given me constant encouragement and although she is on the other side of the country she has never been more than a minute away. I thank my brother for being supportive and for being himself, which has been a wonderful reminder that you should do what makes you happy. And I would especially like to thank my best friend Mike. He has been there for me when I have needed him the most and words cannot express how much I appreciate him.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS iv

LIST OF TABLES ix

LIST OF FIGURES xi

LIST OF SYMBOLS AND ABBREVIATIONS xiv

SUMMARY xvii

CHAPTER

1 INTRODUCTION

V. parahaemolyticus 1

Environmental, Commensal, and Pathogenic Niches of *V. parahaemolyticus* 3

Genetic and Phenotypic Diversity of *V. parahaemolyticus* 6

Pathogenic Mechanisms of *V. parahaemolyticus* 8

Evolution of *V. parahaemolyticus* 13

V. parahaemolyticus as an Emerging Pathogen 20

Dissertation Research Overview 21

References 22

2 MOLECULAR ANALYSIS OF THE EVOLUTION OF *VIBRIO PARAHAEMOLYTICUS* CLINICAL AND ENVIRONMENTAL STRAINS

Objectives and Hypotheses 34

PART 1. MOLECULAR ANALYSIS OF THE EMERGENCE OF PATHOGENIC *VIBRIO PARAHAEMOLYTICUS* STRAINS FROM ENVIRONMENTAL POPULATIONS

Abstract 35

Introduction 36

Materials and Methods 40

Results 47

Discussion 72

PART 2. IDENTIFICATION OF *VIBRIO PARAHAEMOLYTICUS* CLINICAL AND ENVIRONMENTAL STRAINS BY WHOLE-CELL MATRIX ASSISTED LASER DESORPTION IONIZATION-TIME OF FLIGHT ANALYSIS

Abstract 77

Introduction 78

Materials and Methods 81

Results 85

Discussion 108

References 114

3 DIVERSITY OF *VIBRIO* PLASMIDS AND THEIR ROLE IN HORIZONTAL GENE TRANSFER AMONG *VIBRIO PARAHAEMOLYTICUS* STRAINS

Objectives and Hypotheses 120

PART 1. SEQUENCE CHARACTERIZATION AND COMPARATIVE ANALYSIS OF THREE PLASMIDS ISOLATED FROM ENVIRONMENTAL *VIBRIO* SPP.

Abstract 122

Introduction 123

Materials and Methods 124

Results 126

Discussion 126

PART 2. DIVERSITY AND DISTRIBUTION OF *VIBRIO PARAHAEMOLYTICUS* PLASMIDS AND THEIR ROLE IN THE EVOLUTION OF CLINICAL AND ENVIRONMENTAL STRAINS

Abstract 147

Introduction 148

Materials and Methods 151

Results 159

Discussion 182

References 187

4 INACTIVATION OF MISMATCH REPAIR INCREASES THE DIVERSITY OF *V. PARAHAEMOLYTICUS*

Objectives and Hypotheses 196

Abstract 197

Introduction 198

Materials and Methods 200

Results 207

Discussion 224

References 230

5 CONCLUSIONS

Summary of Dissertation Research Findings 235

Broader Implications of Dissertation Research 238

Future Research 242

References 247

VITA 249

LIST OF TABLES

Page

Table 1.1: *V. parahaemolyticus* virulence-associated genes characterized to date 12

Table 1.2: Horizontally transferred *V. parahaemolyticus* genes 14

Table 2.1: Bacterial strains examined 41

Table 2.2: Primers used in the analysis of *V. parahaemolyticus* diversity 43

Table 2.3: Nucleotide diversity of housekeeping and T3SS genes 49

Table 2.4: Virulence gene frequencies 60

Table 2.5: Frequency of MGEs of *V. parahaemolyticus* strains 61

Table 2.6: Swarming motility of *V. parahaemolyticus* strains 62

Table 2.7: Bacterial strains examined 81

Table 2.8: Identification of MALDI peaks 101

Table 2.9: MALDI peak differences of wild-type and mutator strains 108

Table 3.1: Plasmid characteristics 129

Table 3.2: CDSs of plasmid p09022 131

Table 3.3: CDSs of plasmid p0908 136

Table 3.4: CDSs of plasmid p23023 140

Table 3.5: Primers used in the analysis of *V. parahaemolyticus* plasmid diversity 153

Table 3.6: CDSs of plasmid p22702A 162

Table 3.7: CDSs of plasmid p22702B 164

Table 3.8: MGEs of *V. parahaemolyticus* strains 169

Table 4.1: Bacterial strains and vectors used in this study 201

Table 4.2: Primers for the analysis of MMR 203

Table 4.3: *V. parahaemolyticus* MMR homologs 208

Table 4.4: Spontaneous mutation conferring resistance to Rif and Cip 211

Table 4.5: Nucleotide diversity of select wild-type and mutator genes of *V. parahaemolyticus* strains 217

Table 4.6: Transition mutations in *rpoB* of Rif^r colonies 221

Table 4.7: Percentage of translucent (TR) CFU/ml 223

LIST OF FIGURES

Page

Figure 1.1: *V. parahaemolyticus* morphology and motility 3

Figure 1.2: *V. parahaemolyticus* niches 6

Figure 1.3: Pathogenic mechanism of *V. parahaemolyticus* 10

Figure 1.4: Mechanisms of HGT 16

Figure 2.1: Phylogeny of concatenated MLST genes from *V. parahaemolyticus* clinical and environmental strains 52

Figure 2.2: Neighbor-joining phylogeny of *recA* from *V. parahaemolyticus* clinical and environmental strains 54

Figure 2.3: Neighbor-joining phylogeny of *gyrB* from *V. parahaemolyticus* clinical and environmental strains 55

Figure 2.4: Neighbor-joining phylogeny of *pyrC* from *V. parahaemolyticus* clinical and environmental strains 56

Figure 2.5: Neighbor-joining phylogeny of *dtdS* from *V. parahaemolyticus* clinical and environmental strains 57

Figure 2.6: PFGE analysis of *V. parahaemolyticus* strains 58

Figure 2.7: TEM analysis of swarming motility of *V. parahaemolyticus* clonal and non-clonal O3:K6 strains 63

Figure 2.8: Phylogenetic analysis and phenotypic data of *V. parahaemolyticus* clinical and environmental strains 66

Figure 2.9A: Phylogenetic analysis of T3SS1 effector protein-encoding genes 68

Figure 2.9B: Phylogenetic analysis of T3SS1 effector protein-encoding genes 69

Figure 2.10: Phylogenetic analysis of T3SS2 effector protein-encoding genes 71

Figure 2.11: Reproducibility of whole-cell MALDI analysis of *V. parahaemolyticus* RIMD2210633 86

Figure 2.12: MALDI spectra of diverse *Vibrios* 88

Figure 2.13: MALDI spectra of *V. parahaemolyticus* O3:K6 strains 91

Figure 2.14: MALDI spectra of *V. parahaemolyticus* O4:K12 strains 93

Figure 2.15: MALDI spectra of *V. parahaemolyticus* environmental strains 94

Figure 2.16: Cluster analysis of MALDI patterns generated from *Vibrios* 96

Figure 2.17: Dendrogram and PFGE patterns of *V. parahaemolyticus* strains 97

Figure 2.18: Neighbor-joining analysis of concatenated MLST loci of *V. parahaemolyticus* strains 99

Figure 2.19: Whole-cell MALDI spectra of *V. parahaemolyticus* wild-type and mutant strains 106

Figure 3.1: Neighbor-joining analysis of housekeeping genes of the plasmid-bearing *Vibrios* examined 127

Figure 3.2: Genetic organization of plasmid p0908 compared to the Enterobacterial phage P1 133

Figure 3.3: Neighbor-joining analysis of the plasmid-encoded *recA* of p23023 142

Figure 3.4: Sequence alignment of the plasmid-encoded *recA* of p23023 143

Figure 3.5: Genetic organization of the prophage p22702A compared to filamentous vibriophage 161

Figure 3.6: Genetic organization of p22702B compared to the pandemic *V. parahaemolyticus* genomic island VPaI-6 163

Figure 3.7: Sequence alignment of the VPaI-6 gene *vpa1263* compared to the plasmid-encoded gene CDSB13 166

Figure 3.8: Neighbor-joining analysis of housekeeping genes compared to the plasmid replication protein-encoding gene *repA* 171

Figure 3.9: Southern hybridization of a p22702B *rep* probe to other *Vibrio* plasmids 172

Figure 3.10: *V. parahaemolyticus* plasmid restriction endonuclease analysis 175

Figure 3.11: Neighbor-joining analysis of the filamentous phage replication protein-encoding gene *rstA* 178

Figure 3.12: RT-PCR analysis of the expression of *rpoB* and *vpa1263* 180

Figure 3.13: Competition of *V. parahaemolyticus* RIMD2210633 and Δ*vpa1263* 181

Figure 4.1: Genetic organization of the *mutS-rpoS* region of *V. parahaemolyticus* 209

Figure 4.2: Neighbor-joining analysis of *V. parahaemolyticus* wild-type and mutator housekeeping genes 214

Figure 4.3: Neighbor-joining analysis of the MMR gene *mutS* from *V. parahaemolyticus* wild-type and mutator strains 215

LIST OF SYMBOLS AND ABBREVIATIONS

θ_{ω}	Watterson's theta
π	nucleotide diversity
17802	*V. parahaemolyticus* ATCC 17802
22702	*V. parahaemolyticus* 22702
09022	*V. parahaemolyticus* 09022
0908	*V. parahaemolyticus* 0908
23023	*V. parahaemolyticus* 23023
aa	amino acid
AF91	*V. parahaemolyticus* AF91
bp	base pair
˚C	degrees Celsius
Cip	ciprofloxacin
CDS	protein coding sequence
CDC	Centers for Disease Control and Prevention
CFU	colony-forming unit
DNA	deoxyribonucleic acid
dnaE	DNA polymerase III alpha subunit
dtdS	threonine 3-dehydrogenase
GS-PCR	group specific PCR
gyrB	DNA gyrase subunit B
HGT	horizontal gene transfer
HI	heart infusion
Kan	kanamycin

K_a/K_s	ratio of nonsynonymous to synonymous substitutions
kb	kilobase
LB	Luria-Bertani growth medium
M10	marine growth medium
MALDI-TOF MS	matrix-assisted laser desorption ionization-time of flight mass spectrometry
MGE	mobile genetic element
MLST	multilocus sequence typing
ΔmutS	deletion mutant of the mismatch repair gene *mutS*
MMR	mismatch repair
nt	nucleotide
ORF	open reading frame
OP	opaque colony
ΔopaR	deletion mutant of the quorum sensing regulator *opaR*
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
pntA	transhydrogenase alpha subunit
pyrC	dihydro-orotase
recA	DNA recombinase
Rif	rifampicin
RIMD	*V. parahaemolyticus* RIMD2210633
RNA	ribonucleic acid
rRNA	ribosomal RNA
ΔrpoS	deletion mutant of the stationary phase regulator *rpoS*
RT-PCR	reverse transcription PCR

SSCF	sum of squared lengths of condensed fragments
SSUF	sum of squared lengths of uncondensed fragments
Str	streptomycin sulfate
T3SS1	type III secretion system of chromosome I
T3SS2	type III secretion system of chromosome II
tdh	thermostable direct hemolysin
TCBS	thiosulfate citrate bile salts sucrose media
tl	thermolabile hemolysin
tnaA	tryptophanase
TR	translucent colony
trh	*tdh*-related hemolysin
TSA	tryptone soy agar
TSB	tryptone soy broth
VBNC	viable but non-culturable
Δvpa1263	deletion mutant of the putative bacteriocin *vpa1263*

SUMMARY

Vibrio parahaemolyticus is an opportunistic human pathogen that occurs naturally in a non-pathogenic form in coastal estuarine and marine environments worldwide. Following the acquisition of virulence-associated genes, *V. parahaemolyticus* has emerged as a significant pathogen causing seafood-borne illnesses. The mechanisms and conditions that promote the emergence of disease causing *V. parahaemolyticus* strains are not well understood. In addition, *V. parahaemolyticus* clinical strains isolated from disease-associated samples and environmental strains from sediment, water, and marine organisms have been identified with considerable diversity; however, the evolutionary relationships of disease-causing strains and environmental strains are not known. In the following research, the evolutionary relationships of *V. parahaemolyticus* clinical and environmental strains are examined. In addition, the contribution of genetic elements and molecular mechanisms such as deficiency of DNA repair to the evolution of *V. parahaemolyticus* clinical and environmental strains is shown. Molecular analysis of the evolutionary relationships of *V. parahaemolyticus* clinical and environmental strains demonstrated separate lineages of pathogenic and non-pathogenic strains with the exception of several environmental strains that may represent a reservoir of disease-causing strains in the environment. Sequence characterization of plasmids isolated from diverse environmental *Vibrios* indicated a role of plasmids in strain evolution by horizontal transfer of housekeeping genes. In addition, analysis of plasmids from *V. parahaemolyticus* clinical and environmental strains indicated the existence of a plasmid family distributed among *V. parahaemolyticus, V. campbellii,* and *V. harveyi*

environmental strains. Sequence characterization of a plasmid of this family from a *V. parahaemolyticus* environmental strain indicated the contribution of these plasmids to the emergence of the clonal pandemic strains. Investigation of the role of molecular mechanisms to the evolution of *V. parahaemolyticus* strains showed that inactivation of the DNA repair pathway methyl-directed mismatch repair (MMR) increased the accumulation of spontaneous mutations leading to increased nucleotide diversity in select genes. The research findings in the following chapters demonstrate a considerable contribution of genetic elements and molecular mechanisms to the evolution of genetic and phenotypic diversity.

CHAPTER 1

INTRODUCTION

Vibrio parahaemolyticus is a halophilic Gram-negative γ-proteobacterium that occupies diverse aquatic environments and is a significant causative agent of seafood-related illnesses. Of the 63 *Vibrio* species identified to date, *V. parahaemolyticus* is one of 12 *Vibrios* known to cause disease in humans (142). The pathogenic *Vibrios* have been estimated to cause approximately 8,000 illnesses annually in the United States; however, fewer than 10% of the estimated cases are reported (89). *Vibrio*-related foodborne illnesses in the United States have been increasing since 1998 (1). The number of reported *Vibrio*-related infections in the United States in 2006 was 78% higher than those reported during 1996-1998 (1). The majority of *Vibrio*-related human illnesses that occur each year are caused by *V. cholerae, V. vulnificus,* and *V. parahaemolyticus* (142). Of the *Vibrio*-related illnesses reported in the United States in 2006, 64% were associated with *V. parahaemolyticus* (1). Surveillance of *V. parahaemolyticus* infections worldwide is limited; however, *V. parahaemolyticus* is a significant seafood-borne pathogen in Asia (137). As a result, extensive research leading to the isolation and characterization of *V. parahaemolyticus* and identification of its virulence-associated genes has been conducted in Japan (59).

V. parahaemolyticus was first isolated in Japan in 1950 following a disease outbreak due to the consumption of raw fish (45, 46, 98). *V. parahaemolyticus* infections most frequently result from eating raw or undercooked fish and shellfish (1, 3, 15, 30, 142). *V. parahaemolyticus* typically causes gastroenteritis; however, several cases of necrotizing fasciitis have been attributed to *V. parahaemolyticus* (61, 77, 118). Severe infections occur more frequently in immunocompromised individuals and can result in septicemia (2, 3). Although *V. parahaemolyticus* infections are most often contracted as

a foodborne illness, infections have also been reported after recreational water use (2). During the reporting period of 2005-2006 there were 189 *Vibrio*-associated waterborne illnesses in the United States including gastrointestinal, skin, and respiratory infections, with *V. parahaemolyticus* accounting for 33% of the illnesses (2).

An increase in the number of *V. parahaemolyticus*-related infections following the emergence of the O3:K6 clonal pandemic strains has caused *V. parahaemolyticus* to become a pathogen of greater concern (22, 98). The *V. parahaemolyticus* clonal pandemic O3:K6 strains were initially identified in 1995 in India (98) and have since emerged elsewhere in Asia (107, 150), North America (30, 34), South America (48), Europe (111, 125), and Africa (4). In the decade following the emergence of the pandemic clone and related serovars, there has been a striking increase in *Vibrio*-associated illnesses in the United States (1, 30). The majority of the illnesses in the United States are attributed to the O4:K12 serovariants of the O3:K6 pandemic clone (3, 98).

Early research on *V. parahaemolyticus* focused on characterizing hemolytic activity, cell morphology, motility, metabolic properties, and survival. *V. parahaemolyticus* cells are curved rods that have either a single polar flagellum or multiple lateral flagella (84, 86) (Figure 1.1). The single polar flagellum is used for swimming motility while the lateral flagella confer swarming motility that allows rapid colonization of surfaces (67, 84, 86, 135). In addition, extensive research has been conducted on the physiology and survival characteristics of *V. parahaemolyticus.* *Vibrio* spp. have been shown to survive extended starvation and other environmental stressors including changes in temperature and salinity (104, 105, 113, 119, 136). Following periods of nutrient stress and low temperature, *Vibrio* spp. enter into a viable but non-culturable state (VBNC) from which they can be resuscitated (7, 109) facilitating their ability to tolerate changes in temperature and availability of nutrients.

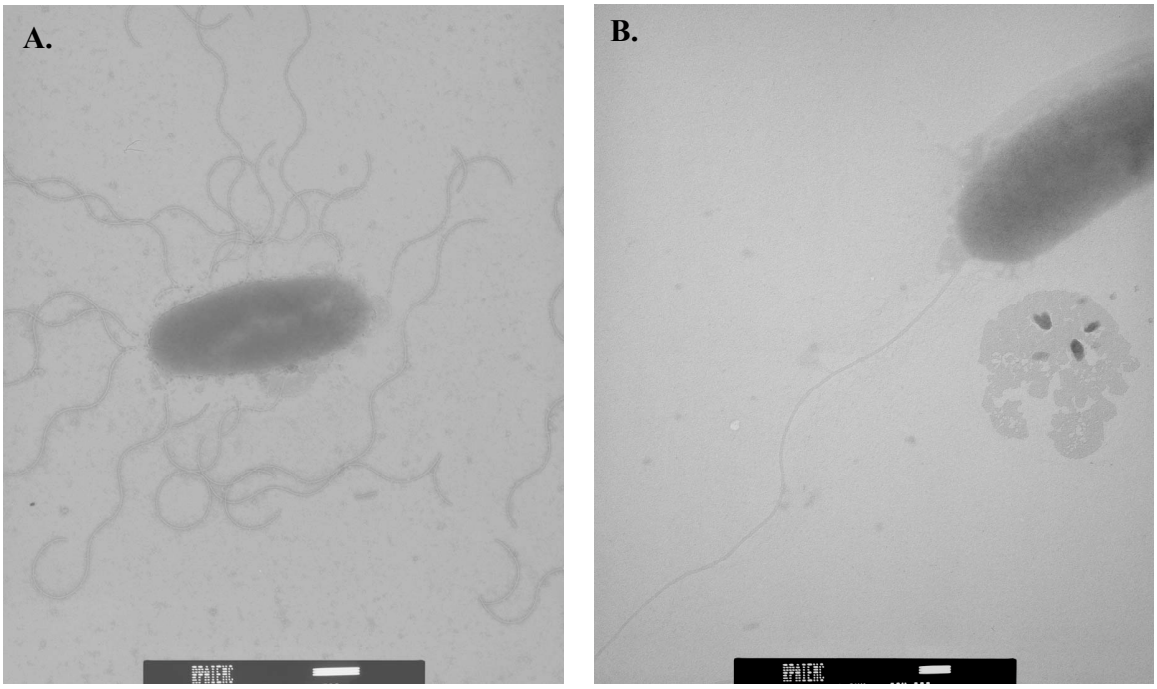


Figure 1.1 TEM images of the *V. parahaemolyticus* cells exhibiting multiple lateral flagella (A) and a cell exhibiting a single polar flagellum (B). The scale bars represent 500 nm (A) and 200 nm (B). The bacteria were grown overnight on media that permits swarming and were negatively stained.

Environmental, commensal, and pathogenic niches of *V. parahaemolyticus*

V. parahaemolyticus occurs in coastal environments in tropical to temperate regions worldwide. *V. parahaemolyticus* has been the causative agent of disease outbreaks as far north as Alaska (88) and as far south as Chile (48). *V. parahaemolyticus* is frequently detected in water, sediment, and organisms from coastal marine environments (16, 27, 32, 33, 159). The wide range of *V. parahaemolyticus* hosts that have been identified to date are shown in Figure 1.2. *V. parahaemolyticus* occurs in estuarine or coastal marine systems as free-living cells or as an assemblage of cells creating a biofilm on the surface of particles or marine organisms such as crustaceans

(142). *V. parahaemolyticus* is frequently isolated from gastropods such as oysters and clams (32, 33).

V. parahaemolyticus and other *Vibrios* adhere to chitinous surfaces and have been isolated from organisms such as crabs, amphipods, copepods, and shrimp (15, 31, 70). *V. parahaemolyticus* has also been isolated from fish (16), although other *Vibrios* such as *V. anguillarum* are more frequently associated with fish diseases (144). *V. parahaemolyticus* was more abundant on the surface of seaweeds than in water in Japan, indicating marine plants may serve as another reservoir for *V. parahaemolyticus* in the environment (78). In addition, the protozoan *Acanthamoeba castellanii* secrete factors that promote the survival of *V. parahaemolyticus* (75).

Marine mammals such as dolphins have been identified as hosts of *Vibrios* and have been documented with *Vibrio*-related illnesses (131). *V. parahaemolyticus* was isolated from wild bottlenose dolphins captured in the Atlantic Ocean and Gulf of Mexico (13). *V. parahaemolyticus* is not known to infect marine mammals; however, the closely-related *V. alginolyticus* caused an infection in a bottlenose dolphin (131). In addition, *V. parahaemolyticus* has been isolated from aquatic birds in Japan indicating that diverse hosts may serve as a reservoir and a vector for dissemination of *V. parahaemolyticus* strains in the environment (96).

The nature of the relationship of *V. parahaemolyticus* with many of its host organisms remains relatively unknown. *Vibrios* cause infections in shrimp that lead to significant losses in revenue from shrimp aquaculture (147). *V. harveyi* and *V. alginolyticus* are the most frequent causative agents of *Vibrio*-associated shrimp diseases; however, *V. parahaemolyticus* has been associated with diseases of juvenile and adult shrimp (147). *V. fischeri* is one of a few symbiotic *Vibrios* that establishes a mutualistic interaction with a host organism (87). *V. logei* and several species of *Photobacterium* are capable of establishing symbioses with squid or fish (142). *V. fischeri* is recruited to the

light organ of the squid *Euprymna scolopes* and produces bioluminescence that camouflages the squid from predators (103).

V. parahaemolyticus reaches high densities in oysters particularly in warmer months (32, 33). The presence of pathogenic and non-pathogenic *V. parahaemolyticus* strains in oysters has been determined using molecular methods such as multiplex and real time-PCR that target the virulence-associated thermostable direct hemolysins and the thermolabile hemolysin that has been identified among all *V. parahaemolyticus* strains (71, 114). Environmental stress such as variable salinity, temperature, and nutrients influence the density of *V. parahaemolyticus* in oysters (102). *V. parahaemolyticus* present in water and sediment can enter bivalves such as oysters during filter-feeding (123). *Vibrios* that enter the oyster gut have been shown to resist being destroyed by haemocytes in the oyster haemolymph (123). *Vibrios* that are able to resist haemocyte killing are presumably then able to colonize oysters tissues and grow to high cell densities (123). *Vibrios* cause disease in mollusc larvae (74); however, the specific contribution of *V. parahaemolyticus* to oyster disease and the impacts on oyster fisheries or aquaculture is unknown (123).

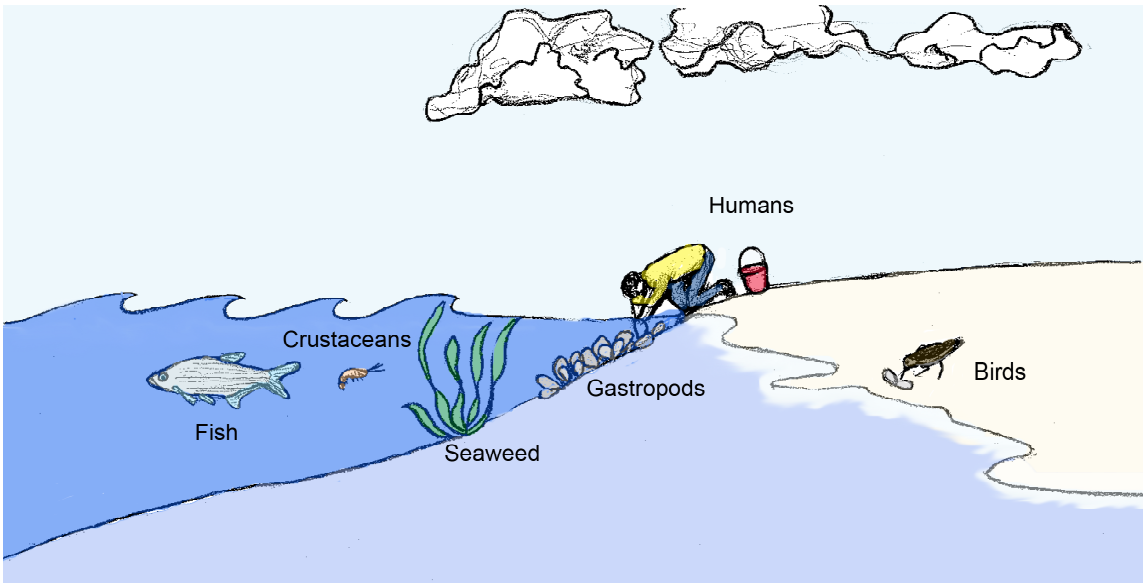


Figure 1.2. *V. parahaemolyticus* niches: *V. parahaemolyticus* is present in sediment and water as free-living planktonic cells or attached biofilms. *V. parahaemolyticus* has been isolated from seaweed (78), birds (96), and dolphins (13) (not shown). In addition, *V. parahaemolyticus* has been isolated from fish, crustaceans, and oysters (16, 32) that are consumed raw or undercooked and cause disease in humans (98).

Genetic and phenotypic diversity

Vibrios have considerable genetic diversity and exhibit relatedness relative to their environmental niche (63). Identification of *V. parahaemolyticus* involves biochemical tests including the selective media thiosulfate citrate bile salts sucrose (TCBS) (40). The standard molecular identification of using 16S rRNA gene analysis (43) is unreliable for distinguishing *V. parahaemolyticus* from *V. harveyi* and *V. campbellii* (142). *V. parahaemolyticus* is closely related to *V. harveyi* and *V. campbellii* based on 16S rRNA gene analysis in addition to the analysis of housekeeping genes (141). *V. parahaemolyticus* has 11 16S rRNA genes that have been shown to undergo recombination (50, 51). Due to a difficulty in identifying *V. parahaemolyticus* by 16S rRNA gene analysis, other housekeeping genes have been examined (23, 139-141, 148).

Among the housekeeping genes analyzed were *recA* and *rpoB* (139, 140, 148). *V. parahaemolyticus* strains isolated from clinical disease-associated samples and strains isolated from environmental samples have been characterized with considerable diversity (49, 81, 82). Due to the genetic and phenotypic relatedness of *V. parahaemolyticus* to other *Vibrios* including *V. harveyi, V. campbellii,* and *V. alginolyticus,* numerous molecular techniques have been employed to identify and characterize the diversity of *V. parahaemolyticus.* Among the frequently used molecular techniques for identification of *V. parahaemolyticus* are pulsed-field gel electrophoresis (PFGE) (68, 81, 117), group-specific PCR (GS-PCR) (9), multiplex PCR (139), multilocus sequence typing (MLST) (23, 49), and gene hybridization (90, 149). The improvement of these molecular techniques or the combination of several techniques has been proven reliable for the accurate identification of disease-causing *V. parahaemolyticus* strains.

Many of the frequently used molecular techniques such as PFGE and MLST are unable to resolve the differences among the clonally-related *V. parahaemolyticus* pandemic strains. The development of new molecular techniques that provide greater resolution would be required to monitor the emergence and spread of *V. parahaemolyticus* clonal pandemic strains. One such method that has shown promise is whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) that has been used for the rapid identification of bacteria such as *Streptococcus pneumoniae* (153)*, Listeria* spp. (5), and *Salmonella* spp. (36). The application of the frequently-used molecular techniques PFGE and MLST are discussed in chapter 2 relative to the development of a whole-cell MALDI approach for identification of *V. parahaemolyticus* (54).

V. parahaemolyticus strains have considerable phenotypic plasticity that reflects their variable environmental or pathogenic niche (82). *V. parahaemolyticus* is a facultative anaerobe and heterotroph capable of using diverse nutrient sources (98). *Vibrios* undergo phenotypic variation in response to changing environmental conditions

such as fluctuating nutrients, temperature, and aeration (57). Phenotypic switching is the change in colony phenotypes that has been observed for *V. vulnificus* environmental strains in response to increased temperature and static growth conditions (57). Similarly, *V. parahaemolyticus* exhibits phenotypic switching resulting in changes in capsular polysaccharide production and in biofilm structure (41, 85). In addition, *V. parahaemolyticus* may horizontally acquire genes that facilitate its niche expansion. *V. natriegens, V. diazotrophicus,* and *V. cincinnatiensis* have previously been shown to fix atmospheric nitrogen (146). We demonstrated in a previous study that a *V. parahaemolyticus* environmental strain isolated from *Spartina*-dominated salt marsh sediment had horizontally-acquired a *nifH* homolog and the ability to fix atmospheric nitrogen at low levels (29).

Pathogenic mechanisms

Hemolysins

V. parahaemolyticus pathogenicity was initially attributed to the presence of the thermostable direct hemolysin *tdh* and the *tdh*-related hemolysin *trh* (58, 60, 101, 130). The thermostable direct hemolysins *tdh* and *trh* cause lysis of red blood cells and are characterized by the detection of ß-hemolysis on blood agar called the Kanagawa phenomenon (24, 95). In addition to hemolysis, several studies have demonstrated that *tdh* is involved in cytotoxicity, enterotoxicity, and cardiotoxicity (100, 101); however, recent research indicates additional genes may be involved in the cytotoxicity and enterotoxicity caused by *V. parahaemolyticus* (8, 14, 116). Previously, *V. parahaemolyticus* isolated from disease events were shown to have one or both of the thermostable direct hemolysins (24). Recent studies have indicated that other mechanisms may be involved in *V. parahaemolyticus* pathogenicity following the isolation of strains from disease-associated samples that did not possess either of the

hemolysins (9, 90). Among the other mechanisms involved in *V. parahaemolyticus* pathogenicity are two type III secretion systems (8, 14, 110, 116).

Type III secretion

Following the completion of the genome sequence of the *V. parahaemolyticus* O3:K6 pandemic strain RIMD2210633, bioinformatic analysis has been used to identify several genomic regions that may be involved in pathogenicity (12, 65). There were 24 genomic regions that were shown to be unique to the *V. parahaemolyticus* pandemic strains (12) indicating acquisition of these regions may have lead to the emergence of the pandemic strains. Among the genomic regions identified that may be involved in pathogenicity of *V. parahaemolyticus* were two type III secretion systems, T3SS1 and T3SS2, located on chromosomes I and II, respectively (79). Type III secretion systems have been shown to be involved in the pathogenic mechanisms of numerous bacteria including *Yersinia* spp., *Pseudomonas aeruginosa,* enteropathogenic (EPEC) and enterohaemorhagic (EHEC) *Escherichia coli,* and *Salmonella* spp. (28). The *V. parahaemolyticus* T3SS1 most resembles the T3SS of *Yersinia* spp. while T3SS2 is most similar to type III of non-O1/non-O139 *V. cholerae* (39, 116). In general, the type III secretion mechanism involves the production of an injectisome that translocates bacterial effector proteins across the bacterial membrane and the eukaryotic membrane into the cytoplasm of a eukaryotic cell (28).

In addition to the T3SS genes, there was a type VI secretion system identified in the genome of the pandemic strain that indicated this region may be involved in the pathogenicity of the pandemic strains (12). Relatively little is known regarding type VI secretion; however, a recent study characterized T6SS genes of *V. cholerae* (124). A summary of all proposed virulence-associated mechanisms associated with pandemic and non-pandemic *V. parahaemolyticus* strains is shown in Figure 1.3.

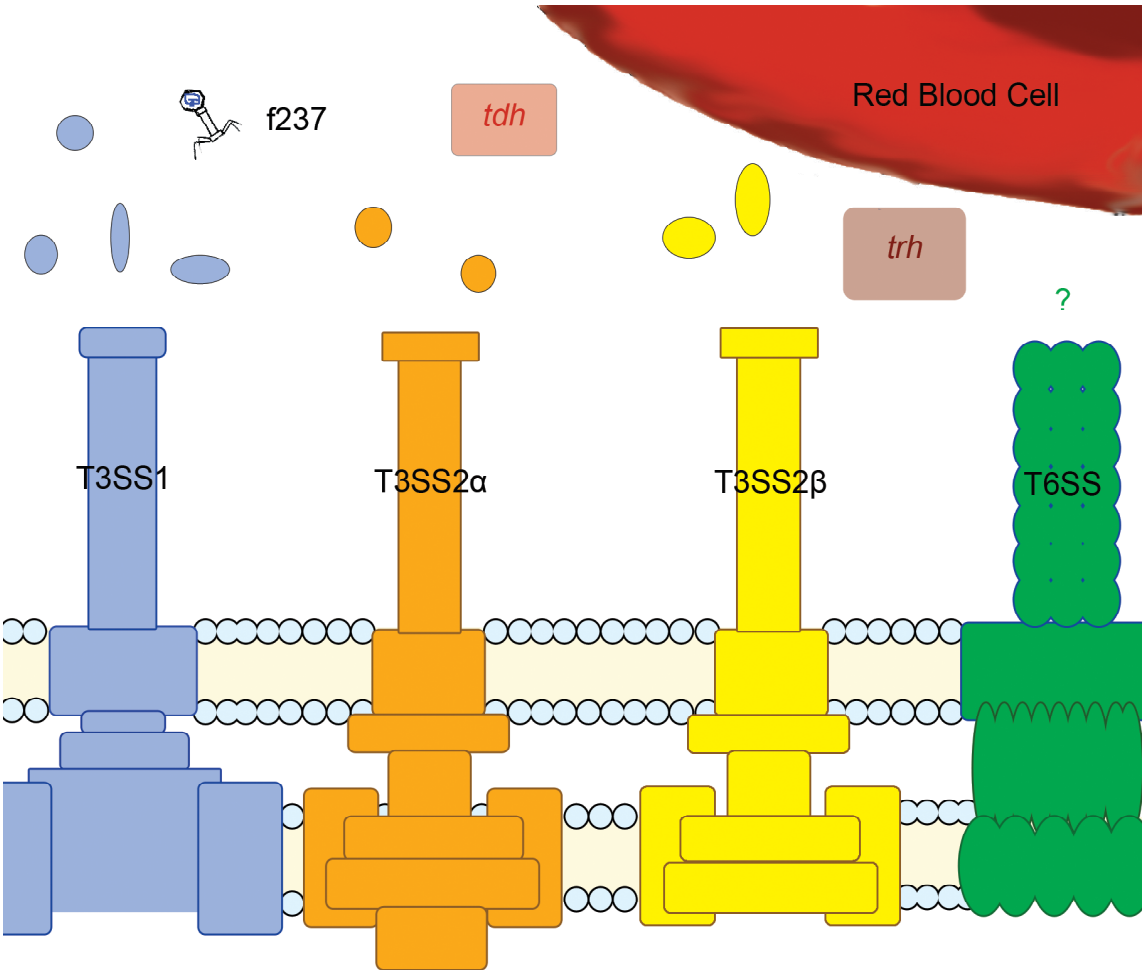


Figure 1.3. A composite illustration of the pathogenic mechanisms of *V. parahaemolyticus*. From top left: filamentous phage f237 containing ORF8 is associated with pandemic strains, the thermostable direct hemolysin *tdh* is associated with pandemic and non-pandemic strains (90), the thermostable direct-related hemolysin *trh* is present in some but not all non-pandemic strains (90), the type VI secretion system (T6SS) is associated with pandemic strains (12), the type III secretion system on chromosome II (T3SS2) present in *trh*-containing non-pandemic strains is T3SS2β (106), the T3SS2 associated with *tdh* in pandemic strains is T3SS2α (65, 79), and the type III secretion system located on chromosome I (T3SS1) is present in all *V. parahaemolyticus* strains (79).

Following the initial genome sequencing of the *V. parahaemolyticus* pandemic strain RIMD2210633 in 2003 (79), several studies have examined the contribution of both of the type III secretion systems to the pathogenic mechanism of *V. parahaemolyticus.* The distribution of all known virulence-associated genes among O3:K6 and non-O3:K6 strains of clinical origin compared to environmental strains isolated from water and sediment samples is shown in Table 1. The summary in Table 1 represents previously published studies as well as my recent research findings discussed in chapter 2.

Investigation of the distribution of these two T3SSs showed that T3SS1 is present among all *V. parahaemolyticus* strains, while T3SS2 is unique to *V. parahaemolyticus* pandemic strains (116). Investigation of the function of T3SS1-secreted proteins revealed they are involved in cytotoxicity toward eukaryotic cells (8, 14, 110, 116). Destruction of eukaryotic cells results from the induction of autophagy, which is the sequestration of cytoplasm in vesicles destroyed by the cellular lysosomes ultimately leading to cell lysis (14).

Bioinformatics analysis indicated that the second type III secretion system (T3SS2) located on the second chromosome and two copies of *tdh* compose a genomic island flanked by Tn7-like transposase genes (12, 138). Investigation of the role of T3SS2 genes for *V. parahaemolyticus* pathogenicity demonstrated that T3SS2 is involved in enterotoxicity (116). The putative pathogenicity island encoding *tdh* and T3SS2 has been demonstrated by sequencing and comparative genome hybridization to be associated with *V. parahaemolyticus* clonal pandemic strains (12, 66). A recent study has demonstrated that the enterotoxicity of pathogenic strains possessing only *trh* and lacking T3SS2 genes can be attributed to the presence of an additional T3SS2 encoded within the region surrounding *trh* (106). This newly-identified T3SS2 is designated T3SS2β and is

located within a 100-kb genomic island present on chromosome II of *V. parahaemolyticus* strains that possess *trh* and not *tdh* (106).

Table 1.1. *V. parahaemolyticus* virulence-associated genes that have been characterized to date

	Serotypes	*tl*	ORF8	*tdh*	*trh*	T3SS1	T3SS2
Clinical	clonal O3:K6	+	+	+	-	+	+
	clonal non-O3:K6	+	+	+	+	+	-
	diverse	+	+	+	+	+	+/-
Environmental	diverse	+	-	+/-	+/-	+	-

The *V. parahaemolyticus* pandemic strains have enhanced swarming compared to non-pandemic strains (155). Swarming motility is the movement of cells across a surface and involves the production of multiple lateral flagella (67). Multiple swarming regulators have been identified for *V. parahaemolyticus* (67, 115). *V. parahaemolyticus* is able to produce either a single polar flagellum or multiple lateral flagella in response to environmental signals (67). Iron limitation and the expression of lateral flagella have been shown to regulate swarming motility (67). In addition, a horizontally acquired gene that encodes an H-NS-like protein, *vpaH*, was shown to regulate the production of lateral flagella in *trh*$^{+}$ *V. parahaemolyticus* strains (115). In other bacteria such as *Xylella fastidiosa*, both iron limitation and iron excess regulate the production of colicin V-like bacteriocins (156).

A putative bacteriocin was identified in the genomic island VPaI-6, which is associated with *V. parahaemolyticus* pandemic strains (65). The role of bacteriocins for virulence and survival of *V. parahaemolyticus* is not known. Among the *Vibrios* that

produce a bacteriocin-like substance is *V. harveyi* (62, 83). A *V. harveyi* strain from an environmental sample produced a bacteriocin-like substance associated with plasmid DNA (83) indicating there may be bacteriocinogenic-like plasmids among *Vibrios*. The *V. harveyi* strain producing the bacteriocin-like substance competed an isogenic plasmid-free strain that no longer produced the bacteriocin (62). *V. vulnificus* also produced a bacteriocin-like substance that inhibited the growth of diverse *Vibrios* including *V. parahaemolyticus* (133). Likewise, a *V. mediterranei* strain was identified that inhibited the growth of *V. parahaemolyticus* (17). In chapter 4, I report the sequence of a plasmid isolated from a *V. parahaemolyticus* environmental strain that had a gene with 98% identity to the putative bacteriocin of the VPaI-6 genomic island (52).

Evolution of *V. parahaemolyticus*

Horizontal gene transfer (HGT)

The completion of nine genome sequences from six *Vibrio* spp. has significantly increased our knowledge of the role of horizontal gene transfer (HGT) in the evolution of *Vibrios* (127). Among the genomes that have been completed are those of the *Vibrio* pathogens *V. cholerae* (56), *V. parahaemolyticus* (79), and *V. vulnificus* (21, 72). In addition, the genome of the fish pathogen *V. splendidus* was recently completed (GenBank) and two genomes of the squid light-organ symbiotic bacterium *V. fischeri* have been sequenced (80, 129). The comparison of multiple *Vibrio* genomes has lead to the identification of numerous genomic regions that have undergone horizontal transfer between *Vibrios* (12, 65). Horizontal transfer of virulence-associated genes has been shown to occur between different *Vibrio* spp. including the acquisition of the *V. parahaemolyticus* thermostable direct hemolysin *trh* by a *V. alginolyticus* strain (47). Comparison of the *V. parahaemolyticus* and *V. cholerae* chromosomes revealed rearrangements between the genomes as well as the acquisition of additional genes on chromosome II of *V. parahaemolyticus* (79). The type III secretion system located on

chromosome II of *V. parahaemolyticus* is most similar to a type III secretion system of non-O1/non-O139 *V. cholerae* strains (39, 106). Other genes shown to be involved in the pathogenic mechanism of *V. parahaemolyticus* that may have been horizontally acquired include *vpaH,* which encodes a histone-like protein involved in swarming motility (115). The effect of horizontal gene transfer on evolution of the genome of *V. parahaemolyticus* has been demonstrated in numerous studies (12, 20, 65, 115). A summary of the *V. parahaemolyticus* genes that were acquired by HGT is presented in Table 1.2.

Table 1.2. Horizontally-acquired genes of *V. parahaemolyticus*

Gene	Size	Protein	Pandemic Strain	MGE	Reference
ORF8	1.4-kb	hypothetical protein	+	Phg	(99)
trh	570-bp	thermostable direct hemolysin	-	Tn	(106)
vpaH	405-bp	H-NS-like protein	-	-	(115)
VPaI-1	24-kb	type I restriction endonuclease, haemagglutinin protein, transcriptional regulator, transmembrane protein, hypothetical proteins	+	Int	(65)
VPaI-2	10-kb	outer membrane protein, resolvase, ribonuclease, hypothetical proteins	+	Int	"
VPaI-3	32-kb	signal transduction histidine kinase, helicase, methyl-accepting chemotaxis protein, AcrBDF protein, hypothetical proteins	+	Int	"
VPaI-4	17-kb	pore-forming cytotoxin integrase, M protein, ATPase, histone deacetylase, hypothetical proteins	+	Int	"
VPaI-5	12-kb	integrases, hypothetical proteins	+	Int	"
VPaI-6	27-kb	DHBP synthase, bacteriocins, hydrolase, hypothetical proteins	+	Int	"
VPaI-7	81-kb	type III secretion system (T3SS2)	+	Tn	(65, 138)
nifH	417-bp	nitrogen fixation	-	-	(29)

The mechanisms of HGT that shape *Vibrio* genomes include transduction, conjugation, and natural transformation (44, 91, 112). *V. cholerae* naturally-transforms free DNA in response to environmental signals that include nutrient limitation and the presence of chitin (91). *V. cholerae* environmental strains transform and recombine DNA segments ranging from 7.9-kb to 44.9-kb (94). The chitin-induced competence mechanism of *V. cholerae* is widely-distributed among *V. cholerae* strains (94). Diverse *Vibrios* including *V. parahaemolyticus* encode the genes required for chitin metabolism (64) and attachment to chitinous surfaces (70, 134); however, to date chitin-induced competence has not been reported for *V. parahaemolyticus* (91, 94). In addition, mobile genetic elements (MGEs) such as phage (20, 99, 151), plasmids (35, 38, 52, 55, 76, 112, 120, 128, 154, 158), and transposable elements (18, 69, 143) are involved in HGT among *Vibrios*. The different mechanisms and barriers of gene transfer are illustrated in Fig. 3. Among the mechanisms that prevent HGT are the production of restriction endonucleases that limit the acquisition of foreign DNA (44). A nuclease from *V. cholerae* was recently shown to lower the frequency of transformation of extracellular DNA (11). In addition, *V. parahaemolyticus* strains have been shown to produce extracellular nucleases (Hazen et al. unpublished). Nucleases serve to protect the cell from the uptake of viral DNA or foreign DNA that is significantly diverged and would likely not undergo homologous recombination (37). Conditions such as nutrient limitation and the presence of chitin have been identified as signals of HGT by stimulating natural transformation in *V. cholerae* (91). Predation by the heteroflagellate *Cafeteria roenbergensis* has also been shown to increase the frequency of conjugal transfer of plasmid DNA among *Vibrio* strains (91).

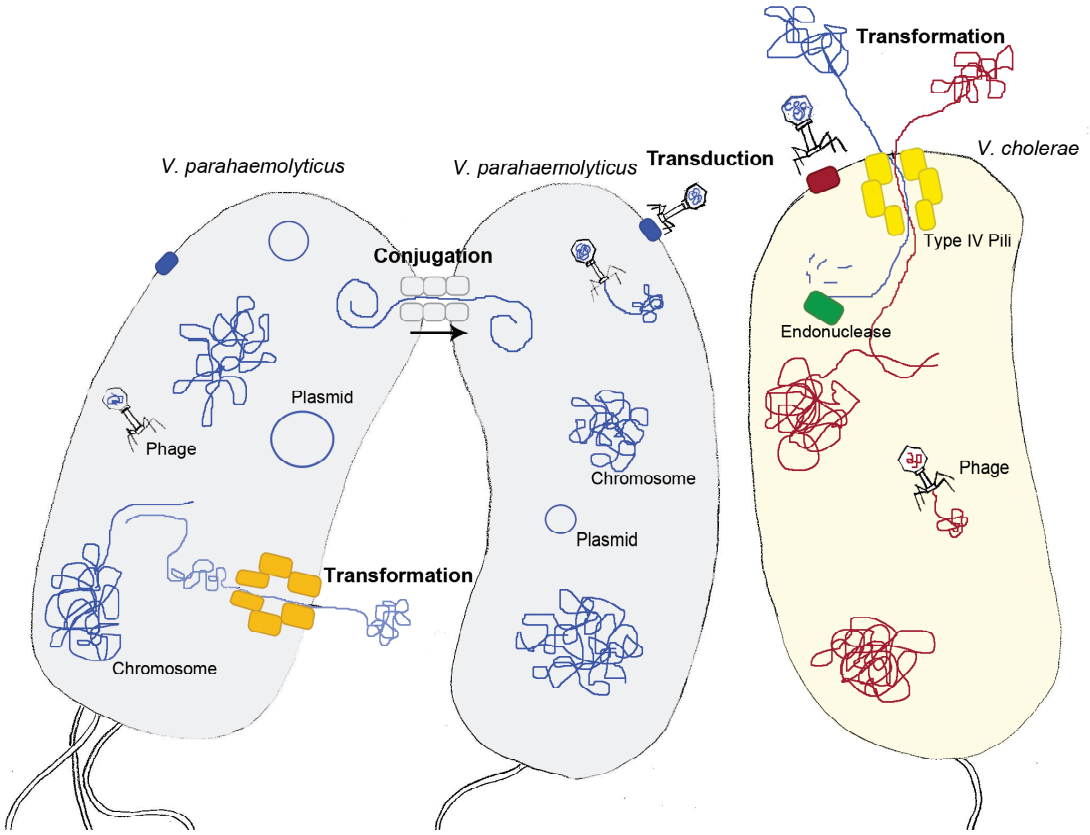


Figure 1.4. Mechanisms and barriers of horizontal gene transfer (HGT) among *Vibrios*. Mechanisms of HGT: conjugation, transduction, and transformation. Barriers to HGT: phage are unable to infect every *Vibrio* host (27), restriction endonucleases and exonucleases released from the cell will prevent the natural transformation of foreign DNA (37).

Mobile genetic elements (MGEs)

Although the role of HGT in shaping *Vibrio* genomes has been clearly demonstrated (127), there is relatively little known regarding the diversity of *V. parahaemolyticus* mobile genetic elements (MGEs) that would serve as the agents of gene transfer. A number of studies have examined the diversity of vibriophage (19, 20, 93, 99, 132); however, few studies have determined the diversity and distribution of plasmids among *V. parahaemolyticus* strains. As discussed in chapter 3, I characterized the DNA sequences of three plasmids isolated from diverse environmental *Vibrios* including *V. mediterranei, V. campbellii,* and *V. fluvialis* (55). Research investigating the presence of bacteriocinogenic *Vibrios* in environmental samples has demonstrated *V. parahaemolyticus* strains have a high frequency of plasmid DNA (133).

Following the discovery of the filamentous phage CTX that encodes the cholera toxin genes *ctxAB* (151), there was significant research conducted on the diversity of filamentous vibriophage (19, 20, 25-27, 99). The filamentous vibriophage f237 was shown to be associated with the *V. parahaemolyticus* clonal pandemic strains (99). Additional vibriophage have been identified among non-pandemic *V. parahaemolyticus* strains that lack ORF8 of f237 (19). The function of ORF8 and its role in the pathogenic mechanism of *V. parahaemolyticus* is unknown; however, ORF8 has been used as a marker for identification of pandemic strains (97). The filamentous phages isolated from *V. parahaemolyticus* have been identified with significant genetic similarity (19, 20, 99). Investigation of the abundance of filamentous vibriophage in environmental samples has shown that vibriophage capable of infecting *V. parahaemolyticus* are abundant in sediment and water (6, 26). In chapter 4, I describe a PCR screen for the most conserved gene of the filamentous phage, the replication protein-encoding gene *rstA*. The results of the PCR screen indicated the phage occurred more frequently among *V. parahaemolyticus* clinical strains compared to environmental strains (52). The sequence of a filamentous phage related to the deleted form of f237 that lacks ORF8 is

characterized in chapter 3 (52). The majority of the *V. parahaemolyticus* phage that have been examined are small (≤10-kb). Several large phages have been characterized that are capable of infecting *V. parahaemolyticus*. Among the large phage is the 244.8-kb T4-like phage KVP40 (93) and the 47.5-kb and 49.5-kb phage VP16C and VP16T, respectively (132).

In comparison to the numerous studies on *V. parahaemolyticus*-associated phage there have been few studies examining the diversity of *V. parahaemolyticus* plasmids. The diversity of virulence-associated plasmids from *V. vulnificus* has been examined with the sequencing of three plasmids that are shown to confer pathogenicity to Biotype 2 strains that infect eels (76). The plasmid encoded *rtx* genes that are known virulence-associated genes, and the loss of the plasmids resulted in a loss of resistance to eel serum and virulence (35). In addition, *V. anguillarum* plasmids have been characterized for their role in pathogenicity toward fish (35). Plasmid pJM1 was shown to encode proteins involved in siderophore production that contribute to *V. anguillarum* pathogenesis (35). In addition, several plasmids have been characterized from the squid symbiont *V. fischeri* including the large plasmid pES100 (45.8-kb) (129) and the small mobilizable plasmid pES213 (5.5-kb) (38). In chapter 3, the DNA sequence of a 28.8-kb plasmid isolated from a *V. parahaemolyticus* environmental strain is characterized (52). Sequence analysis revealed the plasmid had a gene with 98% identity to a putative bacteriocin of the pandemic genomic island VPaI-6 indicating a role of plasmids in horizontal transfer of genes characteristic of pandemic strains (52). In addition, the plasmid had a replication protein-encoding gene (*rep*) with 85% identity to the *rep* of the *V. campbellii* plasmid p09022 that is discussed in chapter 3 (55). Further investigation of the distribution of the conserved *rep*-type revealed a plasmid family present in *V. parahaemolyticus, V. harveyi,* and *V. campbellii* strains isolated from environmental samples (52). Structural analysis by restriction endonuclease digestion of approximately

90-kb plasmids isolated from both *V. parahaemolyticus* clinical and environmental strains are similar (52).

Stress adaptation

Research has examined the occurrence of adaptive mutations for survival in laboratory populations under a controlled type of stress such as nutrient limitation (42). Several molecular mechanisms of stress adaptation characterized for bacteria include induction of the SOS regulon in response to DNA damage, expression of error-prone polymerases, and deficiencies in methyl-directed mismatch repair (MMR) (73). Inactivation of MMR has been shown to occur by mutation of several genes including *mutS* and *mutL*, or by the downregulation of *mutS* by the stationary-phase regulator *rpoS* (145). To date, studies examining the role of inactivation of MMR for adaptive mutation and increased recombination in bacteria have focused on disease-causing bacteria such as *Pseudomonas aeruginosa* (108), *Bacillus anthracis* (157), *Staphylococcus aureus* (121), *Haemophilus influenzae* (152), and *Escherichia coli* (126). Increases in multiple antibiotic resistance have been shown for bacteria involved in chronic infections such as *P. aeruginosa* (108), *H. influenzae* (152), and *S. aureus* (122) from cystic fibrosis patients. Inactivation of MMR and increases in the accumulation of mutation cause pathogenic bacteria responsible for chronic infections to develop resistance to frequently prescribed antibiotics.

There have been no studies to date examining the role of inactivation of MMR for evolution in environmental populations in coastal marine systems. *E. coli* non-pathogenic strains were shown to have increased mutation frequencies during long-term incubations compared to pathogenic strains; however, the contribution of the elevated mutation rate to diversity is not known (10). In chapter 4, I demonstrate that inactivation of the MMR gene *mutS* resulted in an increase in spontaneous mutation of *V. parahaemolyticus* (53). We showed that natural MMR mutants occurred more frequently

among *V. parahaemolyticus* environmental strains compared to clinical strains (53) suggesting a role of inactivation of MMR in the adaptation and survival of this bacterium in the environment. Inactivation of *mutS* in *V. parahaemolyticus* increased phenotypic diversity observed as an increased frequency of phase variation from opaque (OP) to translucent (TR) colonies (53). Deficiencies in MMR may increase the frequency of recombination of DNA segments with increasing sequence divergence as demonstrated previously for *P. stutzeri* (92). An increased frequency of recombination of horizontally-acquired genes would allow HGT of *V. parahaemolyticus* with more distantly-related *Vibrios* such as *V. harveyi* and *V. cholerae*. The role of inactivation of MMR for increases in the recombination of horizontally-acquired DNA is discussed in chapter 4.

V. parahaemolyticus as an emerging pathogen

V. parahaemolyticus outbreaks in colder climates such as Alaska (88) and Chile (48) have indicated an expanding habitat range of *V. parahaemolyticus*. A recent *V. parahaemolyticus*-associated disease outbreak in Alaska coincided with warming water temperatures demonstrating a contribution of global climate change to the occurrence of pathogenic *V. parahaemolyticus* (88). The emergence of the *V. parahaemolyticus* pandemic clone in 1995 and the subsequent increase in the frequency of *V. parahaemolyticus* disease outbreaks (98) has increased the need for research into mechanisms underlying the evolution of this species. In particular, there is extensive evidence of rearrangements and insertions when comparing *Vibrio* genomes as well as the identification of horizontally-transferred virulence-associated genes such as the acquisition of *trh* by *V. alginolyticus* (47). In addition, the emergence of *V. parahaemolyticus* strains with novel virulence characteristics that cause uncommon types of *V. parahaemolyticus* infections such as necrotizing fasciitis that are typical of other *Vibrios* like *V. vulnificus* (98, 118) have increased the need to investigate HGT that occurs among diverse *Vibrio* spp.

Dissertation research overview

Previous research on the diversity of *V. parahaemolyticus* has suggested a significant contribution of HGT to the evolution of disease-causing strains, most notably the emergence of the clonal pandemic strains following HGT of virulence-associated genes (12, 79). The research presented in the following chapters aimed to further characterize the diversity of the pathogenic marine bacterium *Vibrio parahaemolyticus* and the contribution of MGEs and genetic mechanisms to the ongoing evolution of *V. parahaemolyticus.* Research in chapter 2 demonstrates the genetic and phenotypic diversity of *V. parahaemolyticus* disease-causing strains isolated from clinical samples compared to strains isolated from sediment and water samples from coastal environments on the Atlantic coast and Gulf of Mexico in the United States. Differences in pathogenic gene content are characterized for the strains examined and evidence of HGT is discussed. In chapter 3 the diversity of plasmids isolated from diverse environmental *Vibrios* and the role of *V. parahaemolyticus* plasmids in the evolution and emergence of *V. parahaemolyticus* pandemic pathogens is presented. The research objectives of chapter 4 examined the contribution of spontaneous mutation and recombination to the evolution of *V. parahaemolyticus* by determining the role of mismatch repair for increased genetic and phenotypic diversity of clinical compared to environmental strains.

References

1. **(CDC), C. f. D. C. a. P.** 2007. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 states, 2006. MMWR Morb. Mortal. Wkly. Rep. **55:**854-856.
2. **(CDC), C. f. D. C. a. P.** 2008. Surveillance for waterborne disease and outbreaks associated with recreational water use and other aquatic facility-associated health events--- United States, 2005--2006. MMWR Morb. Mortal. Wkly. Rep. **57:**1-29.
3. **(CDC), C. f. D. C. a. P.** 2006. *Vibrio parahaemolyticus* infections associated with consumption of raw shellfish--three states, 2006. MMWR Morb. Mortal. Wkly. Rep. **55:**854-856.
4. **Ansaruzzaman, M., M. Lucas, J. L. Deen, N. A. Bhuiyan, X. Y. Wang, A. Safa, M. Sultana, A. Chowdhury, G. B. Nair, D. A. Sack, L. von Seidlein, M. K. Puri, M. Ali, C. L. Chaignat, J. D. Clemens, and A. Barreto.** 2005. Pandemic serovars (O3:K6 and O4:K68) of *Vibrio parahaemolyticus* associated with diarrhea in Mozambique: spread of the pandemic into the African continent. J. Clin. Microbiol. **43:**2559-2562.
5. **Barbuddhe, S. B., T. Maier, G. Schwarz, M. Kostrzewa, H. Hof, E. Domann, T. Chakraborty, and T. Hain.** 2008. Rapid identification and typing of *Listeria* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. **74:**5402-5407.
6. **Baross, J. A., J. Liston, and R. Y. Morita.** 1978. Incidence of *Vibrio parahaemolyticus* bacteriophages and other *Vibrio* bacteriophages in marine samples. Appl. Environ. Microbiol. **36:**492-499.
7. **Bates, T. C., and J. D. Oliver.** 2004. The viable but nonculturable state of Kanagawa positive and negative strains of *Vibrio parahaemolyticus.* J. Microbiol. **42:**74-79.
8. **Bhattacharjee, R. N., K. S. Park, Y. Kumagai, K. Okada, M. Yamamoto, S. Uematsu, K. Matsui, H. Kumar, T. Kawai, T. Iida, T. Honda, O. Takeuchi, and S. Akira.** 2006. VP1686, a *Vibrio* type III secretion protein, induces toll-like receptor-independent apoptosis in macrophage through NF-kß. J. Biol. Chem. **281:**36897-36904.
9. **Bhoopong, P., P. Palittapongarnpim, R. Pomwised, A. Kiatkittipong, M. Kamruzzaman, Y. Nakaguchi, M. Nishibuchi, M. Ishibashi, and V. Vuddhakul.** 2007. Variability of properties of *Vibrio parahaemolyticus* strains isolated from individual patients. J. Clin. Microbiol. **45:**1544-1550.
10. **Bjedov, I., O. Tenaillon, B. Gérard, V. Souza, E. Denamur, M. Radman, F. Taddei, and I. Matic.** 2003. Stress-induced mutagenesis in bacteria. Science **300:**1404-1409.
11. **Blokesch, M., and G. K. Schoolnik.** 2008. The extracellular nuclease Dns and its role in natural transformation of *Vibrio cholerae*. J. Bacteriol. **190:**7232-7240.
12. **Boyd, E. F., A. L. V. Cohen, L. M. Naughton, D. W. Ussery, T. T. Binnewies, O. C. Stine, and M. A. Parent.** 2008. Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. BMC Microbiol. **8:**110-124.
13. **Buck, J. D., R. S. Wells, H. L. Rhinehart, and L. J. Hansen.** 2006. Aerobic microorganisms associated with free-ranging bottlenose dolphins in coastal Gulf of Mexico and Atlantic Ocean waters. J. Wild. Dis. **42:**536-544.

14. **Burdette, D. L., M. L. Yarbrough, A. Orvedahl, C. J. Gilpin, and K. Orth.** 2008. *Vibrio parahaemolyticus* orchestrates a multifaceted host cell infection by induction of autophagy, cell rounding, and then cell lysis. Proc. Nat. Acad. of Sci. USA **105:**12497-12502.
15. **Cabanillas-Beltrán, H., E. Llausás-Magaña, R. Romero, A. Espinoza, A. García-Gasca, M. Nishibuchi, M. Ishibashi, and B. Gomez-Gil.** 2006. Outbreak of gastroenteritis caused by the pandemic *Vibrio parahaemolyticus* O3:K6 in Mexico. FEMS Microbiol. Lett. **265:**76-80.
16. **Cabrera-García, M. E., C. Vázquez-Salinas, and E. I. Quiñones-Ramírez.** 2004. Serologic and molecular characterization of *Vibrio parahaemolyticus* strains isolated from seawater and fish products of the Gulf of Mexico. Appl. Environ. Microbiol. **70:**6401-6406.
17. **Carraturo, A., K. Raieta, D. Ottaviani, and G. L. Russo.** 2006. Inhibition of *Vibrio parahaemolyticus* by a bacteriocin-like inhibitory substance (BLIS) produced by *Vibrio mediterranei* I. J. Appl. Microbiol. **101:**234-241.
18. **Ceccarelli, D., A. M. Salvia, J. Sami, C. P., and M. M. Colombo.** 2006. New cluster of plasmid-located class 1 integrons in *Vibrio cholerae* O1 and a dfrA15 cassette-containing integron in *Vibrio parahaemolyticus* isolated in Angola. Antimicrob. Agents Chemother. **50:**2493-2499.
19. **Chang, B., H. Miyamoto, H. Taniguchi, and S. Yoshida.** 2002. Isolation and genetic characterization of a novel filamentous bacteriophage, a deleted form of phage f237, from a pandemic *Vibrio parahaemolyticus* O4:K68 strain. Microbiol. Immunol. **46:**565-569.
20. **Chang, B., H. Taniguchi, H. Miyamoto, and S. Yoshida.** 1998. Filamentous bacteriophages of *Vibrio parahaemolyticus* as a possible clue to genetic transmission. J. Bacteriol. **180:**5094-101.
21. **Chen, C. Y., Wu, K. M., Chang, Y. C., Chang, C. H., Tsai, H. C., Liao, T. H., Liu, Y. M., Chen, H. J., Shen, A. B., Li, J. C., Su, T. L., Shao, C. P., Lee, C. T., Hor, L. I., and Tsai, S. F.** 2003. Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. Genome Res. **13:**2577-87.
22. **Chowdhury, N. R., S. Chakraborty, T. Ramamurthy, M. Nishibuchi, S. Yamasaki, Y. Takeda, and G. B. Nair.** 2000. Molecular evidence of clonal *Vibrio parahaemolyticus* pandemic strains. Emerg. Infec. Dis. **6:**631-636.
23. **Chowdhury, N. R., O. C. Stine, J. Glenn Morris, and G. B. Nair.** 2004. Assessment of evolution of pandemic *Vibrio parahaemolyticus* by multilocus sequence typing. J. Bacteriol. **42:**1280-1282.
24. **Chun, D., J. K. Chung, R. Tak, and S. Y. Seol.** 1975. Nature of the Kanagawa phenomenon of *Vibrio parahaemolyticus*. Infect. Immun. **12:**81-87.
25. **Comeau, A. M., E. Buenaventura, and C. A. Suttle.** 2005. A persistent, productive, and seasonally dynamic vibriophage population within Pacific oysters (*Crassostrea gigas*). Appl. Environ. Microbiol. **71:**5324-5331.
26. **Comeau, A. M., A. M. Chan, and C. A. Suttle.** 2006. Genetic richness of vibriophages isolated in a coastal environment. Env. Microbiol. **8:**1164-76.
27. **Comeau, A. M., and C. A. Suttle.** 2007. Distribution, genetic richness, and phage sensitivity of *Vibrio* spp. from coastal British Columbia. Env. Microbiol. **9:**1790-1800.

28. **Cornelis, G. R.** 2006. The type III secretion injectisome. Nat. Rev. Microbiol. **4:**811-825.
29. **Criminger, J. D., T. H. Hazen, P. A. Sobecky, and C. R. Lovell.** 2007. Nitrogen fixation by *Vibrio parahaemolyticus* and its implications for a new ecological niche. Appl. Environ. Microbiol. **73:**5959-5961.
30. **Daniels, N. A., L. MacKinnon, R. Bishop, S. Altekruse, B. Ray, R. M. Hammond, S. Thompson, S. Wilson, N. H. Bean, P. M. Griffin, and L. Slutsker.** 2000. *Vibrio parahaemolyticus* infections in the United States, 1973-1998. J. Infect. Dis. **181:**1661-1666.
31. **Defoirdt, T., R. Crab, T. K. Wood, P. Sorgeloos, W. Verstraete, and P. Bossier.** 2006. Quorum sensing-disrupting brominated furanones protect the gnotobiotic brine shrimp *Artemia franciscana* from pathogenic *Vibrio harveyi, Vibrio campbellii,* and *Vibrio parahaemolyticus* isolates. Appl. Environ. Microbiol. **72:**6419-6423.
32. **DePaola, A., L. H. Hopkins, J. T. Peeler, B. Wentz, and R. M. McPhearson.** 1990. Incidence of *Vibrio parahaemolyticus* in U.S. coastal waters and oysters. Appl. Environ. Microbiol. **56:**2299-2302.
33. **DePaola, A., J. L. Nordstrom, J. C. Bowers, J. G. Wells, and D. W. Cook.** 2003. Seasonal abundance of total and pathogenic *Vibrio parahaemolyticus* in Alabama oysters. Appl. Environ. Microbiol. **69:**1521-1526.
34. **DePaola, A., J. Ulaszek, C. A. Kaysner, B. J. Tenge, J. L. Nordstrom, J. Wells, N. Puhr, and S. M. Gendel.** 2003. Molecular, serological, and virulence characteristics of *Vibrio parahaemolyticus* isolated from environmental, food, and clinical sources in North America and Asia. Appl. Environ. Microbiol. **69:**3999-4005.
35. **Di Lorenzo, M., M. Stork, M. E. Tolmasky, L. A. Actis, D. Farrell, T. J. Welch, L. M. Crosa, A. M. Wertheimer, Q. Chen, P. Salinas, L. Waldbeser, and J. H. Crosa.** 2003. Complete sequence of virulence plasmid pJM1 from the marine fish pathogen *Vibrio anguillarum* strain 775. J. Bacteriol. **185:**5822-5830.
36. **Dieckmann, R., R. Helmuth, M. Erhard, and B. Malorny.** 2008. Rapid classification and identification of Salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. **74:**7767-7778.
37. **Dubnau, D.** 1999. DNA uptake in bacteria. Annu. Rev. Microbiol. **53:**217-244.
38. **Dunn, A. K., M. O. Martin, and E. V. Stabb.** 2005. Characterizatioon of pES213, a small mobilizable plasmid from *Vibrio fischeri*. Plasmid **54:**114-134.
39. **Dziejman, M., D. Serruto, V. C. Tam, D. Sturtevant, P. Diraphat, S. M. Faruque, M. H. Rahman, J. F. Heidelberg, J. Decker, L. Li, K. T. Montgomery, G. Grills, R. Kucherlapati, and J. J. Mekalanos.** 2005. Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. Proc. Natl. Acad. Sci. USA **102:**3465-3470.
40. **Elliot, E. L., C. A. Kaysner, L. Jackson, and M. L. Tamplin.** 1998. *Vibrio cholerae, V. parahaemolyticus, V. vulnificus,* and other *Vibrio* spp., p. 9.01-9.27. *In* U. S. F. a. D. A. B. A. Manual (ed.). A.O.A.C. International, Gaithersburg, MD.

41. **Enos-Berlage, J. L., Z. T. Guvener, C. E. Keenan, and L. L. McCarter.** 2005. Genetic determinants of biofilm development of opaque and translucent *Vibrio parahaemolyticus*. Mol. Microbiol. **55:**1160-1182.
42. **Finkel, S. E., and R. Kolter.** 1999. Evolution of microbial diversity during prolonged starvation. Proc. Nat. Acad. of Sci. USA **96:**4023-4027.
43. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. S. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **209:**457-63.
44. **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint.** 2005. Mobile genetic elements: the agents of open source evolution. Nat. Rev. Microbiol. **3:**722-732.
45. **Fujino, T., T. Miwatani, J. Yasuda, M. Kondo, Y. Takeda, Y. Akita, K. Kotera, M. Okada, H. Nishimune, Y. Shimizu, T. Tamura, and Y. Tamura.** 1965. Taxonomic studies on the bacterial strains isolated from cases of "shirasu" food-poisoning (*Pasteurella parahaemolytica*) and related microorganisms. Biken J. **8:**63-71.
46. **Fujino, T., Y. Okuno, D. Nakada, A. Aoyama, K. Fukai, T. Mukai, and T. Uebo.** 1953. On the bacteriological examination of Shirasu food poisoning. Med. J. Osaka Univ. **4:**299-304.
47. **González-Escalona, N., G. M. Blackstone, and A. DePaola.** 2006. Characterization of a *Vibrio alginolyticus* strain, isolated from Alaskan oysters, carrying a hemolysin gene similar to the thermostable direct hemolysin-related hemolysin gene (*trh*) of *Vibrio parahaemolyticus*. Appl. Environ. Microbiol. **72:**7925-7929.
48. **González-Escalona, N., V. Cachicas, C. Acevedo, M. L. Rioseco, J. A. Vergara, F. C. Cabello, J. Romero, and R. T. Espejo.** 2005. *Vibrio parahaemolyticus* diarrhea, Chile, 1998 and 2004. Emerg. Infec. Dis. **11:**129-131.
49. **González-Escalona, N., J. Martinez-Urtaza, J. Romero, R. T. Espejo, L. Jaykus, and A. DePaola.** 2008. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. J. Bacteriol. **190:**2831-2840.
50. **Gonzalez-Escalona, N., J. Romero, and R. T. Espejo.** 2005. Polymorphism and gene conversion of the 16S rRNA genes in the multiple rRNA operons of *Vibrio parahaemolyticus.* FEMS Microbiol. Lett. **246:**213-219.
51. **Harth, E., J. Romero, R. Torres, and R. T. Espejo.** 2007. Intragenomic heterogeneity and intergenomic recombination among *Vibrio parahaemolyticus* 16S rRNA genes. Microbiology **153:**2640-2647.
52. **Hazen, T. H., D. J. Silberger, D. Wu, Garrett, N., Bopp, C. A., J. A. Eisen, and P. A. Sobecky.** In Submission. Diversity of mobile genetic elements isolated from *Vibrio parahaemolyticus* clinical and environmental strains and the identification of a plasmid-encoded bacteriocin.
53. **Hazen, T. H., K. D. Kennedy, S. Chen, S. V. Yi, and P. A. Sobecky.** 2009. Inactivation of mismatch repair increases the diversity of *Vibrio parahaemolyticus*. Env. Microbiol. **11:**1254-1266.

54. **Hazen, T. H., R. J. Martinez, Y. Chen, P. C. Lafon, N. M. Garrett, M. M. Parsons, C. A. Bopp, M. C. Sullards, and P. A. Sobecky.** In Submission. Identification of *Vibrio parahaemolyticus* by whole-cell matrix assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol.
55. **Hazen, T. H., D. Wu, J. A. Eisen, and P. A. Sobecky.** 2007. Sequence characterization and comparative analysis of three plasmids isolated from environmental *Vibrio* spp. Appl. Environ. Microbiol. **73:**7703-7710.
56. **Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter, and C. M. Fraser.** 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406:**477-83.
57. **Hilton, T., T. Rosche, B. Froelich, B. Smith, and J. Oliver.** 2006. Capsular polysaccharide phase variation in *Vibrio vulnificus*. Appl. Environ. Microbiol. **72:**6986-6993.
58. **Honda, S., I. Goto, I. Minematsu, N. Ikeda, N. Asano, M. Ishibashi, Y. Kinoshita, M. Nishibuchi, T. Honda, and T. Miwatani.** 1987. Gastroenteritis due to Kanagawa negative *Vibrio parahaemolyticus*. Lancet **1:**331-332.
59. **Honda, T., T. Iida, Y. Akeda, and T. Kodama.** 2008. Sixty years of *Vibrio parahaemolyticus* research. Microbe **3:**462-466.
60. **Honda, T., Y. X. Ni, and T. Miwatani.** 1988. Purification and characterization of a hemolysin produced by a clinical isolate of Kanagawa phenomenon-negative *Vibrio parahaemolyticus* and related to the thermostable direct hemolysin. Infect. Immun. **56:**961-965.
61. **Howard, R. J., M. E. Pessa, B. H. Brennaman, and R. Ramphal.** 1985. Necrotizing soft-tissue infections caused by marine *Vibrios*. Surgery **98:**126-130.
62. **Hoyt, P. R., and R. K. Sizemore.** 1982. Competitive dominance by a bacteriocin-producing *Vibrio harveyi* strain. Appl. Environ. Microbiol. **44:**653-658.
63. **Hunt, D. E., L. A. David, D. Gevers, S. P. Preheim, E. J. Alm, and M. F. Polz.** 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science **320:**1081-1085.
64. **Hunt, D. E., D. Gevers, N. M. Vahora, and M. F. Polz.** 2008. Conservation of the chitin utilization pathway in the *Vibrionaceae*. Appl. Environ. Microbiol. **74:**44-51.
65. **Hurley, C. C., A. M. Quirke, F. J. Reen, and E. F. Boyd.** 2006. Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates. BMC Genom. **7:**104.
66. **Izutsu, K., K. Kurokawa, K. Tashiro, S. Kuhara, T. Hayashi, T. Honda, and T. Iida.** 2008. Comparative genomic analysis using microarray demonstrates a strong correlation between the presence of the 80-kilobase pathogenicity island and pathogenicity in Kanagawa phenomenon-positive *Vibrio parahaemolyticus* strains. Infect. Immun. **76:**1016-1023.

67. **Jacques, S., and L. L. McCarter.** 2006. Three new regulators of swarming in *Vibrio parahaemolyticus*. J Bacteriol **188:**2625-2635.
68. **Kam, K. M., C. K. Y. Luey, M. B. Parsons, K. L. F. Cooper, G. B. Nair, M. Alam, M. A. Islam, D. T. L. Cheung, Y. W. Chu, T. Ramamurthy, G. P. Pazhani, S. K. Bhattacharya, H. Watanabe, J. Terajima, E. Arakawa, O.-A. Ratchtrachenchai, S. Huttayananont, E. M. Ribot, P. Gerner-Smidt, and B. Swaminathan.** 2008. Evaluation and validation of a PulseNet standardized pulsed-field gel electrophoresis protocol for subtyping *Vibrio parahaemolyticus:* an international multicenter collaborative study. J. Clin. Microbiol. **46:**2766-2773.
69. **Kamruzzaman, M., and M. Nishibuchi.** 2008. Detection and characterization of a functional insertion sequence, ISVpa2, in *Vibrio parahaemolyticus*. Gene **409:**92-99.
70. **Kaneko, T., and R. R. Colwell.** 1975. Adsorption of *Vibrio parahaemolyticus* onto chitin and copepods. Appl. Microbiol. **29:**269-274.
71. **Kaufman, G. E., G. M. Blackstone, M. C. Vickery, A. K. Bej, J. Bowers, M. D. Bowen, R. F. Meyer, and A. DePaola.** 2004. Real-time PCR quantification of *Vibrio parahaemolyticus* in oysters using an alternative matrix. J. Food Prot. **67:**2424-2429.
72. **Kim, Y. R., S. E. Lee, C. M. Kim, S. Y. Kim, E. K. Shin, D. H. Shin, S. S. Chung, H. E. Choy, A. Progulske-Fox, J. D. Hillman, M. Handfield, and J. H. Rhee.** 2003. Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients. Infect. Immun. **71:**5461-5471.
73. **Kivisaar, M.** 2003. Stationary phase mutagenesis: mechanisms that accelerate adaptation of microbial populations under environmental stress. Environ. Microbiol. **5:**814-827.
74. **Lambert, C., and J. L. Nicolas.** 1998. Specific inhibition of chemiluminescent activity by pathogenic *Vibrios* in hemocytes of two marine bivalves: *Pecten maximus* and *Crassostrea gigas*. J. Invertebr. Pathol. **71:**53-63.
75. **Laskowski-Arce, M. A., and K. Orth.** 2008. *Acanthamoeba castellanii* promotes the survival of *Vibrio parahaemolyticus*. Appl. Environ. Microbiol. **74:**7183-7188.
76. **Lee, C. T., C. Amaro, K. M. Wu, E. Valiente, Y. F. Change, S. F. Tsai, C. H. Chang, and L. I. Hor.** 2008. A common virulence plasmid in biotype 2 *Vibrio vulnificus* and its dissemination aided by a conjugal plasmid. J. Bacteriol. **190:**1638-48.
77. **Lim, T. K., and A. E. Stebbings.** 1999. Fulminant necrotosing fasciitis caused by *Vibrio parahaemolyticus*. Singapore Med. J. **40:**596-597.
78. **Mahmud, Z. H., S. B. Neogi, A. Kassu, T. Wada, M. S. Islam, G. B. Nair, and F. Ota.** 2007. Seaweeds as a reservoir for diverse *Vibrio parahaemolyticus* populations in Japan. Int. J. Food Microbiol. **118:**92-96.
79. **Makino, K., K. Oshima, K. Kurokawa, K. Yokoyama, T. Uda, K. Tagomori, Y. Iijima, M. Najima, M. Nakano, A. Yamashita, Y. Kubota, S. Kimura, T. Yasunaga, T. Honda, H. Shinagawa, M. Hattori, and T. Iida.** 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. Lancet **361:**743-749.

80. **Mandel, M. J., M. S. Wollenberg, E. V. Stabb, K. L. Visick, and E. G. Ruby.** 2009. A single regulatory gene is sufficient to alter bacterial host range. Nature **458:**15-23.
81. **Martinez-Urtaza, J., A. Lozano-Leon, A. DePaola, M. Ishibashi, K. Shimada, M. Nishibuchi, and E. Liebana.** 2004. Characterization of pathogenic *Vibrio parahaemolyticus* isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates. J. Clin. Microbiol. **42:**4672-8.
82. **Martinez-Urtaza, J., A. Lozano-Leon, A. Viña-Feas, J. de Novoa, and O. Garcia-Martin.** 2006. Differences in the API 20E biochemical patterns of clinical and environmental *Vibrio parahaemolyticus* isolates. FEMS Microbiol. Lett. **255:**75-81.
83. **McCall, J. O., and R. K. Sizemore.** 1979. Description of a bacteriocinogenic plasmid in *Beneckea harveyi*. Appl. Environ. Microbiol. **38:**974-9.
84. **McCarter, L. L.** 2004. Dual flagellar systems enable motility under different circumstances. J. Mol. Biol. Tech. **7:**18-29.
85. **McCarter, L. L.** 1998. OpaR, a homolog of *Vibrio harveyi* LuxR, controls opacity of *Vibrio parahaemolyticus*. J. Bacteriol. **180:**3166-3173.
86. **McCarter, L. L.** 2001. Polar flagellar motility of the Vibrionaceae. Microb. Molec. Biol. Rev. **65:**445-462.
87. **McFall-Ngai, M. J., and E. G. Ruby.** 1991. Symbiont recognition and subsequent morphogenesis as early events in an animal-bacterial mutualism. Science **254:**1491-1494.
88. **McLaughlin, J. B., A. DePaola, C. A. Bopp, K. A. Martinek, N. P. Napolilli, C. G. Allison, S. L. Murray, E. C. Thompson, M. M. Bird, and J. P. Middaugh.** 2005. Outbreak of *Vibrio parahaemolyticus* gastroenteritis associated with Alaskan oysters. N. Engl. J. Med. **353:**1463-70.
89. **Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe.** 1999. Food-related illness and death in the United States. Emerg. Infec. Dis. **5:**607-625.
90. **Meador, C. E., M. B. Parsons, C. A. Bopp, P. Gerner-Smidt, J. A. Painter, and G. J. Vora.** 2007. Virulence gene- and pandemic group-specific marker profiling of clinical *Vibrio parahaemolyticus* isolates. J. Clin. Microbiol. **45:**1133-1139.
91. **Meibom, K. L., M. Blokesch, N. A. Dolganov, C. Y. Wu, and G. K. Schoolnik.** 2005. Chitin induces natural competence in *Vibrio cholerae*. Science **310:**1824-1827.
92. **Meier, P., and W. Wackernagel.** 2005. Impact of *mutS* inactivation on foreign DNA acquisition by natural transformation in *Pseudomonas stutzeri*. J. Bacteriol. **187:**143-154.
93. **Miller, E. S., J. F. Heidelberg, J. A. Eisen, W. C. Nelson, A. S. Durkin, A. Ciecko, T. V. Feldblyum, O. White, I. T. Paulsen, W. C. Nierman, J. Lee, B. Szczypinski, and C. M. Fraser.** 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: Comparative genomics of a T4-related bacteriophage. J. Bacteriol. **185:**5220-5233.
94. **Miller, M. C., D. P. Keymer, A. Avelar, A. B. Boehm, and G. K. Schoolnik.** 2007. Detection and transformation of genome segments that differ within a

coastal population of *Vibrio cholerae* strains. Appl. Environ. Microbiol. **73:**3695-7304.

95. **Miyamoto, Y., T. Kato, Y. Obara, S. Akiyama, K. Takizawa, and S. Yamai.** 1969. In vitro hemolytic characteristic of *Vibrio parahaemolyticus*: its close correlation with human pathogenicity. J. Bacteriol. **100:**1147-1149.
96. **Miyasaka, J., S. Yahiro, Y. Arahira, H. Tokunaga, K. Katsuki, and Y. Hara-Kudo.** 2006. Isolation of *Vibrio parahaemolyticus* and *Vibrio vulnificus* from wild aquatic birds in Japan. Epidemiol. Infect. **134:**780-785.
97. **Myers, M. L., G. Panicker, and A. K. Bej.** 2003. PCR detection of a newly emerged pandemic *Vibrio parahaemolyticus* O3:K6 pathogen in pure cultures and seeded waters from the Gulf of Mexico. Appl. Environ. Microbiol. **69:**2194-2200.
98. **Nair, G. B., T. Ramamurthy, S. K. Bhattacharya, B. Dutta, Y. Takeda, and D. A. Sack.** 2007. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. Clin. Microbiol. Rev. **20:**39-48.
99. **Nasu, H., T. Iida, T. Sugahara, Y. Yamaichi, K. S. Park, K. Yokoyama, K. Makino, H. Shinagawa, and T. Honda.** 2000. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. J. Clin. Microbiol. **38:**2156-2161.
100. **Nishibuchi, M., A. Fasano, R. G. Russell, and J. B. Kaper.** 1992. Enterotoxigenicity of *Vibrio parahaemolyticus* with and without genes encoding thermostable direct hemolysin. Infect. Immun. **60:**3539-3545.
101. **Nishibuchi, M., and J. B. Kaper.** 1995. Thermostable direct hemolysin gene of *Vibrio parahaemolyticus*: a virulence gene acquired by a marine bacterium. Infect. Immun. **63:**2093-2099.
102. **Nordstrom, J. L., C. A. Kaysner, G. M. Blackstone, M. C. Vickery, J. C. Bowers, and A. DePaola.** 2004. Effect of intertidal exposure on *Vibrio parahaemolyticus* levels in Pacific Northwest oysters. J. Food. Prot. **67:**2178-2182.
103. **Nyholm, S. V., E. V. Stabb, E. G. Ruby, and M. J. McFall-Ngai.** 2000. Establishment of an animal-bacterial association: recruiting symbiotic *Vibrios* from the environment. Proc. Nat. Acad. of Sci. USA **97:**10231-10235.
104. **Nystrom, T., K. Flardh, and S. Kjelleberg.** 1990. Responses to multiple-nutrient starvation in marine *Vibrio* sp strain-Ccug-15956. J. Bacteriol. **172:**7085-7097.
105. **Nystrom, T., R. M. Olsson, and S. Kjelleberg.** 1992. Survival, stress resistance, and alterations in protein expression in the marine *Vibrio* sp strain S14 during starvation for different individual nutrients. Appl. Environ. Microbiol. **58:**55-65.
106. **Okada, N., T. Iida, K. Park, N. Goto, T. Yasunaga, H. Hiyoshi, S. Matsuda, T. Kodama, and T. Honda.** 2009. Identification and characterization of a novel type III secretion system in *trh*-positive *Vibrio parahaemolyticus* strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level. Infect. Immun. **77:**904-913.
107. **Okuda, J., M. Ishibashi, E. Hayakawa, T. Nishino, Y. Takeda, A. K. Mukhopadhyay, S. Garg, S. K. Bhattacharya, G. B. Nair, and M. Nishibuchi.** 1997. Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in

Calcutta, India, and isolation of strains from the same clonal group from southeast Asian travelers arriving in Japan. J. Clin. Microbiol. **35:**3150-3155.

108. **Oliver, A., R. Cantón, P. Campo, F. Baquero, and J. Blázquez.** 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science **288:**1251-1253.
109. **Oliver, J. D., L. Nilsson, and S. Kjelleberg.** 1991. Formation of nonculturable *Vibrio vulnificus* cells and its relationship to the starvation state. Appl. Environ. Microbiol. **57:**2640-2644.
110. **Ono, T., K. S. Park, M. Ueta, I. Tetsuya, and T. Honda.** 2006. Identification of proteins secreted via *Vibrio parahaemolyticus* Type III secretion system 1. Infect. Immun. **74:**1032-1042.
111. **Ottaviani, D., F. Leoni, E. Rocchegiani, S. Santarelli, C. Canonico, L. Masini, V. Ditrani, and A. Carraturo.** 2008. First clinical report of pandemic *Vibrio parahaemolyticus* O3:K6 infection in Italy. J. Clin. Microbiol. **46:**2144-2145.
112. **Otto, K., D. Weichart, and S. Kjelleberg.** 1997. Plasmid transfer between marine *Vibrio* strains during predation by the heterotrophic microflagellate *Cafeteria roenbergensis*. Appl. Environ. Microbiol. **63:**749-752.
113. **PaludanMuller, C., D. Weichart, D. McDougald, and S. Kjelleberg.** 1996. Analysis of starvation conditions that allow for prolonged culturability of Vibrio vulnificus at low temperature. Microbiology-Uk **142:**1675-1684.
114. **Panicker, G., D. R. Call, M. J. Krug, and A. K. Bej.** 2004. Detection of pathogenic *Vibrio* spp. in shellfish by using multiplex PCR and DNA microarrays. Appl. Environ. Microbiol. **70:**7436-7444.
115. **Park, K. S., M. Arita, T. Iida, and T. Honda.** 2005. *vpaH,* a gene encoding a novel histone-like nucleoid structure-like protein that was possibly horizontally acquired, regulates the biogenesis of lateral flagella in *trh*-positive *Vibrio parahaemolyticus* TH3996. Infect. Immun. **73:**5754-5761.
116. **Park, K. S., T. Ono, M. Rokuda, M. H. Jang, K. Okada, T. Iida, and T. Honda.** 2004. Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. Infect. Immun. **72:**6659-6665.
117. **Parsons, M. B., K. L. Cooper, K. A. Kubota, N. Puhr, S. Simington, P. S. Calimlim, D. Schoonmaker-Bopp, C. Bopp, B. Swaminathan, P. Gerner-Smidt, and E. M. Ribot.** 2007. PulseNet USA standardized pulsed-field gel electrophoresis protocol for subtyping of *Vibrio parahaemolyticus*. Foodborne Pathog. Dis. **4:**285-92.
118. **Payinda, G.** 2008. Necrotizing fasciitis due to *Vibrio parahaemolyticus*. N. Z. Med. J. **121:**99-101.
119. **Pérez-Rosas, N., and T. C. Hazen.** 1989. In situ survival of *Vibrio cholerae* and *Escherichia coli* in a tropical rain forest watershed. Appl. Environ. Microbiol. **55:**495-499.
120. **Powers, L. G., J. T. Mallonee, and P. A. Sobecky.** 2000. Complete nucleotide sequence of a cryptic plasmid from the marine bacterium *Vibrio splendidus* and identification of open reading frames. Plasmid **43:**99-102.
121. **Prunier, A. L., and R. Leclercq.** 2005. Role of *mutS* and *mutL* genes in hypermutability and recombination in *Staphylococcus aureus*. J. Bacteriol. **187:**3455-3464.

122. **Prunier, A. L., B. Malbruny, M. Laurans, J. Brouard, J. F. Duhamel, and R. Leclercq.** 2003. High rate of macrolide resistance in *Staphylococcus aureus* strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. J. Infect. Dis. **187:**1709-1716.
123. **Pruzzo, C., G. Gallo, and L. Canesi.** 2005. Persistence of *Vibrios* in marine bivalves: the role of interactions with haemolymph components. Environ. Microbiol. **7:**761-772.
124. **Pukatzki, S., A. T. Ma, A. T. Revel, D. Sturtevant, and J. J. Mekalanos.** 2007. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. Proc. Nat. Acad. of Sci. USA **104:**15508-15513.
125. **Quilici, M. L., A. Robert-Pillot, J. Picart, and J. M. Fournier.** 2005. Pandemic *Vibrio parahaemolyticus* O3:K6 spread, France. Emerg. Infec. Dis. **11:**1148-1149.
126. **Rayssiguier, C., Thaler, D. S., and Radman, M.** 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature **342:**396-401.
127. **Reen, F. J., S. Almagro-Moreno, D. Ussery, and E. F. Boyd.** 2006. The genomic code: inferring *Vibrionaceae* niche specialization. Nat. Rev. Microbiol. **9:**697-704.
128. **Rubin, E. J., W. Lin, J. J. Mekalanos, and M. K. Waldor.** 1998. Replication and integration of a *Vibrio cholerae* cryptic plasmid linked to the CTX prophage. Mol. Microbiol. **28:**1247-1254.
129. **Ruby, E. G., M. Urbanowski, J. Campbell, A. Dunn, M. Faini, R. Gunsalus, P. Lostroh, C. Lupp, J. McCann, D. Millikan, A. Schaefer, E. Stabb, A. Stevens, K. Visick, C. Whistler, and E. P. Greenberg.** 2005. Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners. Proc. Natl. Acad. Sci. USA **102:**3004-3009.
130. **Sakurai, J., A. Matsuzaki, Y. Takeda, and T. Miwatani.** 1974. Existence of two distinct hemolysins in *Vibrio parahaemolyticus*. Infect. Immun. **9:**777-780.
131. **Schroeder, J. P., J. G. Wallace, M. B. Cates, S. B. Greco, and P. W. Moore.** 1985. An infection by *Vibrio alginolyticus* in an Atlantic bottlenose dolphin housed in an open ocean pen. J. Wild. Dis. **21:**437-438.
132. **Seguritan, V., I. W. Feng, F. Rohwer, M. Swift, and A. M. Segall.** 2003. Genome sequences of two closely related *Vibrio parahaemolyticus* phages, VP16T and VP16C. J. Bacteriol. **185:**6434-6447.
133. **Shehane, S. D., and R. K. Sizemore.** 2002. Isolation and preliminary characterization of bacteriocins produced by *Vibrio vulnificus*. J. Appl. Microbiol. **92:**322-8.
134. **Shime-Hattori, A., T. Iida, M. Arita, K. Park, T. Kodama, and T. Honda.** 2006. Two type IV pili of *Vibrio parahaemolyticus* play different roles in biofilm formation. FEMS Microbiol. Lett. **264:**89-97.
135. **Stewart, B. J., and L. L. MCCarter.** 2003. Lateral flagellar gene system of *Vibrio parahaemolyticus*. J. Bacteriol. **185:**4508-4518.
136. **Stretton, S., S. J. Danon, S. Kjelleberg, and A. E. Goodman.** 1997. Changes in cell morphology and motility in the marine Vibrio sp Strain S14 during conditions of starvation and recovery. Fems Microbiology Letters **146:**23-29.

137. **Su, Y. C., and C. Liu.** 2007. *Vibrio parahaemolyticus*: a concern of seafood safety. Food Microbiol. **24:**549-558.
138. **Sugiyama, T., T. Iida, K. Izutsu, K. S. Park, and T. Honda.** 2008. Precise region and the character of the pathogenicity island in clinical *Vibrio parahaemolyticus* strains. J. Bacteriol. **190:**1835-1837.
139. **Tarr, C. L., J. S. Patel, N. D. Puhr, E. G. Sowers, C. A. Bopp, and N. A. Strockbine.** 2007. Identification of *Vibrio* isolates by a multiplex PCR assay and *rpoB* sequence determination. J. Clin. Microbiol. **45:**134-40.
140. **Thompson, C. C., F. L. Thompson, K. Vandemeulebroecke, B. Hoste, P. Dawyndt, and J. Swings.** 2004. Use of *recA* as an alternative phylogenetic marker in the family *Vibrionaceae*. Int. J. Syst. Evol. Microbiol. **54:**919-924.
141. **Thompson, F. L., D. Gevers, C. C. Thompson, P. Dawyndt, S. Naser, B. Hoste, C. B. Munn, and J. Swings.** 2005. Phylogeny and molecular identification of *Vibrios* on the basis of multilocus sequence analysis. Appl. Environ. Microbiol. **71:**5107-5115.
142. **Thompson, F. L., T. Iida, and J. Swings.** 2004. Biodiversity of *Vibrios*. Microbiol. Mol. Biol. Rev. **68:**403-431.
143. **Tolmasky, M. E., and J. H. Crosa.** 1995. Iron transport genes of the pJM1-mediated iron uptake system of *Vibrio anguillarum* are included in a transposonlike structure. Plasmid **33:**180-190.
144. **Toranzo, A. E., Y. Santos, and J. L. Barja.** 1997. Immunization with bacterial antigens: *Vibrio* infections. Dev. Biol. Stand. **90:**93-105.
145. **Tsui, H. C., G. Feng, and M. E. Winkler.** 1997. Negative regulation of *mutS* and *mutH* repair gene expression by the Hfq and RpoS global regulators of *Escherichia coli* K-12. J. Bacteriol. **179:**7476-7487.
146. **Urdaci, M. C., L. J. Stal, and M. Marchand.** 2004. Occurrence of nitrogen fixation among *Vibrio* spp. Arch. Microbiol. **150:**224-229.
147. **Vandenberghe, J., L. Verdonck, R. Robles-Arozarena, G. Rivera, A. Bolland, M. Balladares, B. Gomez-Gil, J. Calderon, P. Sorgeloos, and J. Swings.** 1999. *Vibrios* associated with *Litopenaeus vannamei* larvae, postlarvae, broodstock, and hatchery probionts. Appl. Environ. Microbiol. **65:**2592-2597.
148. **Venkateswaran, K., N. Dohmoto, and S. Harayama.** 1998. Cloning and nucleotide sequence of the *gyrB* gene of *Vibrio parahaemolyticus* and its application in detection of this pathogen in shrimp. Appl. Environ. Microbiol. **64:**681-687.
149. **Vora, G. J., C. E. Meador, M. M. Bird, C. A. Bopp, J. D. Andreadis, and D. A. Stenger.** 2005. Microarray-based detection of genetic heterogeneity, antimicrobial resistance, and the viable but nonculturable state in human pathogenic *Vibrio* spp. Proc. Natl. Acad. Sci. USA **102:**19109-19114.
150. **Vuddhakul, V., A. Chowdhury, V. Laohaprertthisan, P. Pungrasamee, N. Patararungrong, P. Thianmontri, M. Ishibashi, C. Matsumoto, and M. Nishibuchi.** 2000. Isolation of a pandemic O3:K6 clone of a *Vibrio parahaemolyticus* strain from environmental and clinical sources in Thailand. Appl. Environ. Microbiol. **66:**2685-2689.
151. **Waldor, M. K., and J. J. Mekalanos.** 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science **272:**1910-1914.

152. **Watson, M. E., J. L. Burns, and A. L. Smith.** 2004. Hypermutable *Haemophilus influenzae* with mutations in *mutS* are found in cystic fibrosis sputum. Microbiology-Sgm **150:**2947-2958.
153. **Williamson, Y. M., H. Moura, A. R. Woolfitt, J. L. Pirkle, J. R. Barr, M. D. G. Carvalho, E. P. Ades, G. M. Carlone, and J. S. Sampson.** 2008. Differentiation of *Streptococcus pneumoniae* conjunctivitis outbreak isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. **74:**5891-5897.
154. **Wu, H., Y. Ma, Y. Zhang, and H. Zhang.** 2004. Complete sequence of virulence plasmid pEIB1 from the marine fish pathogen *Vibrio anguillarum* strain MVM425 and location of its replication region. J. Appl. Microbiol. **97:**1021-1028.
155. **Yeung, P. S. M., M. C. Hayes, A. DePaola, C. A. Kaysner, L. Kornstein, and K. J. Boor.** 2002. Comparative phenotypic, molecular, and virulence characterization of *Vibrio parahaemolyticus* O3:K6 isolates. Appl. Environ. Microbiol. **68:**2901-2909.
156. **Zaini, P. A., A. C. Fogaca, F. G. N. Lupo, H. I. Nakaya, R. Z. N. Vêncio, and A. M. da Silva.** 2008. The iron stimulon of *Xylella fastidiosa* includes genes for type IV pilus and colicin V-like bacteriocins. J. Bacteriol. **190:**2368-2378.
157. **Zeibell, K., S. Aguila, V. Y. Shi, A. Chan, H. Yang, and J. H. Miller.** 2007. Mutagenesis and repair in *Bacillus anthracis*: the effect of mutators. J. Bacteriol. **189:**2331-2338.
158. **Zhang, R., Y. Wang, P. C. Leung, and J. D. Gu.** 2007. pVC, a small cryptic plasmid from the environmental isolate of *Vibrio cholerae* MP-1. J. Microbiol. **45:**193-198.
159. **Zimmerman, A. M., A. DePaola, J. C. Bowers, J. A. Krantz, J. L. Nordstrom, C. N. Johnson, and D. J. Grimes.** 2007. Variability of total and pathogenic *Vibrio parahaemolyticus* densities in northern Gulf of Mexico water and oysters. Appl. Environ. Microbiol. **73:**7589-7596.

CHAPTER 2

MOLECULAR ANALYSIS OF THE EVOLUTION OF *VIBRIO PARAHAEMOLYTICUS* CLINICAL AND ENVIRONMENTAL STRAINS

The objective of the research described in chapter 2 was to determine the evolutionary relationship of *V. parahaemolyticus* clinical and environmental strains and whether there are reservoirs of disease-causing strains in the environment. I hypothesized that *V. parahaemolyticus* environmental strains are an underestimated reservoir of intermediate pathogens and emerging pandemic clones. The objectives of part 1 of chapter 2 were to: 1) use molecular techniques to examine the genetic diversity of *V. parahaemolyticus* strains isolated from coastal regions in the southeastern United States, 2) compare the genetic diversity of environmental strains to clinical strains associated with illnesses in the U.S., and 3) determine the virulence-associated gene content of the *V. parahaemolyticus* strains and examine the nucleotide diversity of the type III secretion system effector protein-encoding genes.

The second section of chapter 2 discusses the development of a method for the rapid identification of *V. parahaemolyticus* clinical and environmental strains. The objectives of this study were: 1) to determine whether whole-cell MALDI-TOF MS analysis is an effective method for identification of *V. parahaemolyticus* strains, and 2) examine whether MALDI-TOF MS analysis provides greater resolution for distinguishing among clonal *V. parahaemolyticus* strains compared with other molecular techniques. I discuss the identification of *V. parahaemolyticus* strains by whole-cell matrix assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) analysis compared to conventional molecular techniques frequently used

including multilocus sequence typing (MLST) and pulsed-field gel electrophoresis (PFGE).

Molecular Analysis of the Emergence of Pathogenic *Vibrio parahaemolyticus* Strains From Environmental Populations

Abstract

Vibrio parahaemolyticus is an opportunistic disease-causing marine bacterium that is prevalent in coastal environments; however, few of the naturally occurring strains that have been characterized possess the known virulence-associated genes. The mechanism of *V. parahaemolyticus* pathogenicity has not been fully characterized, although several genes have been linked to cytotoxicity and enterotoxicity of eukaryotic cells. Previously, the thermostable direct hemolysins (*tdh/trh*) were considered the main virulence mechanism of *V. parahaemolyticus*; however, recent studies have demonstrated the contribution of two type III secretion systems (T3SS1 and T3SS2) to virulence. Previous studies that analyzed the virulence reservoir of *V. parahaemolyticus* in the environment by detecting the hemolysins may have underestimated the emerging pathogen potential of *V. parahaemolyticus* environmental populations. In this study, I analyzed the genetic structure and virulence gene content of a total of 85 *V. parahaemolyticus* strains consisting of 28 clinical strains isolated from disease-associated samples and 57 environmental strains isolated from coastal sediment and water of the southeastern United States. Multilocus sequence analysis of four housekeeping genes demonstrated that *V. parahaemolyticus* clinical strains formed a separate group from the environmental strains. There were ten strains isolated from environmental samples that

grouped with the clinical strains, including a strain isolated from sediment in FL that was similar to the O3:K6 pandemic clones.

Introduction

Vibrio parahaemolyticus is a leading cause of seafoodborne gastroenteritis in the United States and is one of 12 *Vibrio* spp. that have been shown to cause illness in humans (17, 70). Infections caused by *V. parahaemolyticus* are typically gastroenteritis, but less frequently wound infections and septicemia (48). The pathogenic mechanisms of *V. parahaemolyticus* have not been completely characterized; however, several virulence-associated genes including two thermostable direct hemolysins and two type III secretion systems that have been linked to enterotoxicity and cytotoxicity to eukaryotic cells (11, 57). Although pathogenicity of *V. parahaemolyticus* has been linked to hemolysins and type III secretion, disease-causing strains have been identified with diverse virulence gene profiles indicating that virulence may have evolved multiple times and likely involves genes that have yet to be characterized (7, 46).

An increase in the number of *V. parahaemolyticus*-associated infections occurred following the emergence and spread of a clonal pandemic strain with the O3:K6 serotype (48, 52). The O3:K6 clonal pandemic strain was first characterized in 1995 from a disease outbreak in India (52), and has since been isolated elsewhere in Asia, Europe, Africa, North America, and South America (2, 12, 47, 48, 54, 60, 72). Recently, disease-causing strains have been characterized with the additional serotypes O1:KUT, O1:K25, and O4:K68 indicating the clonal strains have diverged into these serotypes (48). In the United States, one of the serotypes associated with many of the illnesses each year is

O4:K12, which is also clonal and may be derived from the O3:K6 pandemic strains (1, 46, 48).

Methods for examining the population diversity and evolution of *V. parahaemolyticus* clinical strains have employed multilocus sequence typing (MLST), microarrays and macroarrays (35, 46, 71), and genome subtraction (53). The analysis of *V. parahaemolyticus* clinical strains by genome subtraction was used to identify an 80-kb pathogenicity island (Vp-PAI) associated with the clonal pandemic strains (53). In addition, bioinformatic analysis of the *V. parahaemolyticus* O3:K6 pandemic strain RIMD2210633 genome revealed seven genomic islands (34). Of the genomic islands, VPaI-7/Vp-PAI was shown to be specific to the pandemic strains and encoded a type III secretion system and two copies of thermostable direct hemolysin (*tdh*) (34). Multilocus sequence typing of housekeeping genes showed that there were two clonal complexes in addition to the pandemic clonal complex (26). One of the newly-identified clonal complexes consisted of O4:K12 and O12:K12 strains that were isolated from the Pacific coast of the United States (26).

The thermostable direct hemolysins were identified as virulence-associated genes for their role in β-hemolysis of red blood cells (15). More recent studies have demonstrated the role of two type III secretion systems (T3SS1 and T3SS2) located on chromosomes I and II for cytotoxicity and enterotoxicity, respectively (43, 57). A second T3SS2 (T3SS2β) was recently identified in *trh*-positive strains that lacked *tdh* (51). The T3SS2β genes of *trh* strains were identified in a 100-kb region surrounding *trh* on chromosome II (51). In addition, a filamentous phage was identified among pandemic *V. parahaemolyticus* strains indicating a gene unique to this phage may be involved in the

increased pathogenicity associated with the pandemic strains (49). The filamentous phage identified among pandemic strains was identical to other filamentous phage isolated from non-O3:K6 *V. parahaemolyticus* pandemic strains; however, the pandemic O3:K6 phage possessed an additional protein-encoding gene, ORF8 (13, 49). Due to its association with the pandemic *V. parahaemolyticus* O3:K6 strains, ORF8 has been identified as a potential virulence marker; however, the function of this gene is unknown (49) and virulence has been primarily attributed to the chromosomally-encoded thermostable direct hemolysins (57, 63).

In a previous study, T3SS1 genes were identified among all *V. parahaemolyticus* strains examined while T3SS2 was identified only among Kanagawa-positive (KP) strains (43). The Kanagawa phenomenon (KP) is the detection of β-hemolysis of red blood cells caused by *tdh*. Bioinformatic analysis of the *V. parahaemolyticus* genome indicated T3SS2 and two copies of *tdh* are located within the VPaI-7 genomic island (34). Recent studies have shown that *V. parahaemolyticus* clinical strains possess the *tdh* gene but none of the T3SS2 genes indicating the *tdh* genes are capable of spontaneously deleting from the genomic island (7, 46). In fact, the *tdh* is associated with insertion sequences that facilitate spontaneous deletion and acquisition of this gene (37).

Identification of *tdh* on plasmid DNA isolated from *V. parahaemolyticus* and *V. cholerae* and the characterization of the filamentous phage of pandemic strains has indicated a role of MGEs in the horizontal transfer of virulence-associated genes. In addition, T3SS1 that is present in all *V. parahaemolyticus* strains is most related to a type III secretion system of *V. harveyi,* while T3SS2 is similar to a type III secretion system of non-O1/non-O139 *V. cholerae* strains (21, 57). These findings suggest that T3SS1 may

have been acquired by *V. parahaemolyticus* and *V. harveyi* from a common ancestor, while the T3SS2 genes have likely been horizontally-acquired and subsequently lost in certain strains (9).

To date, the primary genes that are detected as a confirmation of *V. parahaemolyticus* pathogenicity are *tdh* and *trh* and few studies have examined the distribution of the T3SS genes (34, 43, 57). In the current study we examined the frequency of T3SS1 and T3SS2 genes among a greater number of environmental *Vibrio* strains and analyzed the diversity of two effector protein-encoding genes secreted by each of the T3SSs. We show that all the *V. parahaemolyticus* environmental strains examined possessed T3SS1 and only the O3:K6 clonal strains had T3SS2. In addition, we used MLST to determine the population structure of geographically co-occurring and separated *V. parahaemolyticus* environmental strains from coastal regions of the southeastern United States. Phylogenetic analysis revealed several *V. parahaemolyticus* clinical and environmental strains that may be in an intermediate stage of pathogenicity acquisition or loss.

Materials and Methods

Bacterial strains and media. The *V. parahaemolyticus* clinical strains were isolated from samples associated with human illness and were provided by the Centers for Disease Control and Prevention (Atlanta, GA). Environmental strains were isolated from sediment, water, and oysters of Skidaway Island, GA and Apalachicola Bay, FL in September 2006 as previously described (30). Additional strains were isolated from the rhizosphere sediment of a salt marsh in North Inlet, NC and were provided by C. Lovell (4). Environmental strains examined in this study that formed green colonies on thiosulfate citrate bile salts sucrose (TCBS) agar were confirmed to be *V. parahaemolyticus* by screening for the thermolabile hemolysin (*tl*) that is characteristic of *V. parahaemolyticus* (46). The presumptive *V. parahaemolyticus* strains that had grown as green colonies on TCBS agar were screened for MGE content as described below and an approximately equal number of strains with and without MGEs were selected for further analysis. A total of 85 *V. parahaemolyticus* strains were examined for their population structure and diversity of virulence genes. Of the strains examined there were 57 strains from environmental samples and 28 strains from clinical samples (Table 2.1).

Table 2.1. *V. parahaemolyticus* clinical and environmental strains examined

Strain Id	Year	Location	Serotype	Source	MGEs	*tl*	ORF8	*tdh*	*trh*	VP1670	VP1680	VP1686	VPA1346	VPA1354	VPA1362
RIMD2210633	1996	Japan	O3:K6	clinical	+	+	+	+	-	+	+	+	+	+	+
17802	1965	Japan	O3:K6	clinical	-	+	-	+	-	+	+	+	-	-	-
K1533	2004	CT	O3:K6	clinical	-	+	+	+	-	+	+	+	+	+	+
K1223	2004	CO	O3:K6	clinical	+	+	+	+	-	+	+	+	+	+	+
K4435	2006	GA	O3:K6	clinical	-	+	+	+	-	+	+	+	+	+	+
F9083	2002	AZ	O3:K6	clinical	-	+	+	+	-	+	+	+	+	+	+
F8186	2001	NY	O4:K12	clinical	-	+	-	+	+	+	+	+	-	-	-
F8023	2001	GA	O4:K12	clinical	-	+	-	+	+	+	+	+	-	-	-
K1461	2004	MA	O4:K12	clinical	+	+	-	+	+	+	+	+	-	-	-
K5281	2007	WA	O4:K12	clinical	-	+	-	+	+	+	+	+	-	-	-
F5052	1997	WA	O4:K12	clinical	-	+	+	+	+	+	+	+	-	-	-
F7044	2000	TX	O4:K12	clinical	-	+	-	+	+	+	+	+	-	-	-
K4250	2006	NY	O4:K63	clinical	+	+	-	+	+	+	+	+	-	-	-
K3566	2006	LA	O4:K63	clinical	-	+	-	+	+	+	+	+	-	-	-
K0851	2004	LA	O1:Kuk	clinical	-	+	-	-	-	+	+	+	-	-	-
K0081	2003	AZ	O6:K18	clinical	+	+	-	+	+	+	+	+	-	-	-
K0850	2004	LA	O1:K25	clinical	-	+	-	-	-	+	+	+	-	-	-
F7979	2001	MN	O5:K17	clinical	+	+	+	-	-	+	+	+	-	-	-
F6179	1998	CT	O6:K18	clinical	+	+	-	-	-	+	-	+	-	-	-
F8937	2002	NY	O4:K10	clinical	+	+	-	-	-	+	+	+	+	+	+
K4237	2006	NY	O10:K56	clinical	+	+	-	-	-	+	+	+	-	-	-
K4376	2006	MD	O8:K21	clinical	+	+	-	+	+	+	+	+	-	-	-
K4434	2006	MS	O6:K18	clinical	-	+	-	-	-	+	+	+	-	-	-
K4638	2007	NY	O10:Kuk	clinical	-	+	-	-	-	+	+	+	-	-	-
K4981	2007	OK	O1:Kuk	clinical	-	+	+	-	-	+	+	+	-	-	-
K5067	2007	SD	O1:K56	clinical	+	+	-	+	+	+	+	+	-	-	-
K5323 g	2007	VA	O5:K17	clinical	-	+	-	-	-	+	+	+	-	-	-
K5330	2007	TX	O5:Kuk	clinical	-	+	-	-	-	+	+	+	-	-	-
22702	1998	GA	O5:Kuk	sediment	+	+	-	-	-	+	+	+	-	-	-
SG58	2006	GA	ND	water	-	+	-	-	-	+	+	+	-	-	-
SG151	2006	GA	ND	water	-	+	-	-	-	+	+	+	-	-	-
SG158	2006	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
SG174	2006	GA	ND	water	-	+	-	-	-	+	+	+	-	-	-
SG176	2006	GA	O5:Kuk	water	+	+	-	-	-	+	+	+	-	-	-
SG209	2006	GA	O2:K28	sediment	+	+	-	-	-	+	+	+	-	-	-
SG214	2006	GA	O1:K9	water	+	+	-	-	-	+	+	+	-	-	-
SG227	2006	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
SG244	2006	GA	O6:Kuk	water	+	+	-	-	-	+	+	+	-	-	-
SG247	2006	GA	ND	water	-	+	-	-	-	+	+	+	-	-	-
SG255	2006	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
SG257	2006	GA	O8:Kuk	sediment	+	+	-	-	+	+	+	+	-	-	-
SG258	2006	GA	O1:Kuk	sediment	+	+	-	-	-	+	+	+	-	-	-
SG259	2006	GA	ND	sediment	+	+	-	-	+	+	+	+	-	-	-
SG285	2006	GA	O10:Kuk	sediment	+	+	-	-	-	+	+	+	-	-	-
SG328	2006	GA	ND	water	+	+	-	-	-	+	+	+	-	-	-
SG354	2006	GA	O2:K28	water	+	+	-	-	-	+	+	+	-	-	-
SG364	2006	GA	O1:Kuk	sediment	+	+	-	-	-	+	+	+	-	-	-
SG372	2006	GA	O5:Kuk	sediment	+	+	-	-	-	+	+	+	-	-	-
SG398	2006	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
SG401	2006	GA	O5:K61	sediment	+	+	-	-	-	+	+	+	-	-	-
SG409	2006	GA	O1:K33	sediment	+	+	-	-	-	+	+	+	-	-	-
SG417	2006	GA	O4:K67	sediment	+	+	-	-	-	+	+	+	-	-	-
AF2	2006	FL	O4:K9	oyster	+	+	-	-	-	+	+	+	-	-	-
AF8	2006	FL	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
AF12	2006	FL	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
AF34	2006	FL	O5:K37	sediment	+	+	-	-	-	+	+	+	-	-	-
AF46	2006	FL	O11:Kuk	water	+	+	-	-	-	+	+	+	-	-	-
AF47	2006	FL	ND	water	+	+	-	-	-	+	+	+	-	-	-
AF56	2006	FL	O6:K37	sediment	+	+	-	-	-	+	+	+	-	-	-
AF64	2006	FL	ND	sediment	+	+	-	-	-	+	+	+	-	-	-
AF67	2006	FL	O5:K37	sediment	+	+	-	-	-	+	+	+	-	-	-
AF91	2006	FL	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
AF97	2006	FL	ND	water	-	+	-	-	-	+	+	+	-	-	-
J-C2-6	1998	NC	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
J-C2-15	1998	NC	O1:Kuk	sediment	+	+	-	-	-	+	+	+	-	-	-
J-C2-29	1998	NC	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
J-C2-34	1998	NC	O5:K19	sediment	+	+	-	-	-	+	+	+	-	-	-
J-M1-23	1998	NC	O1:K32	sediment	+	+	-	-	-	+	+	+	-	-	-
S-M2-2-B3	1998	NC	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
S-M2-3-B3	1998	NC	O5:K17	sediment	+	+	-	-	-	+	+	+	-	-	-
7-SG1	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
7-SG12	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
7-SG38	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
7-SG62	2007	GA	ND	sediment	+	+	-	-	-	+	+	+	-	-	-
7-SG81	2007	GA	ND	sediment	+	+	-	-	-	+	+	+	-	-	-
7-SG84	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
7-SG127	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
7-SG146	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
7-SG159	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
7-SG165	2007	GA	ND	sediment	+	+	-	-	-	+	+	+	-	-	-
7-SG167	2007	GA	ND	sediment	+	+	-	-	-	+	+	+	-	-	-
7-SG170	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
7-SG180	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-
7-SG204	2007	GA	ND	sediment	+	+	-	-	-	+	+	+	-	-	-
7-SG206	2007	GA	ND	sediment	-	+	-	-	-	+	+	+	-	-	-

PCR detection of virulence-associated genes. The known virulence-associated genes were detected by PCR for all clinical and environmental strains examined in this study. We confirmed that strains were *V. parahaemolyticus* by detecting the thermolabile hemolysin (*tl*) using primers previously developed (6). All strains that were positive for *tl* were then PCR screened for previously characterized virulence-associated genes. We screened for the presence of the thermostable direct hemolysins *tdh* and *trh* as previously described (6, 46). In addition, we detected the ORF8 gene of the pandemic phage f237 using primers that were previously developed (47). To determine the presence of T3SS1 and T3SS2, we PCR screened for two effectors and one structural gene from each T3SS. From T3SS1 we amplified the effectors *vp1680* and *vp1686* and the structural gene *vp1670* using primers listed in Table 2.2. To confirm the presence of T3SS2 we screened for the effectors *vpa1346* and *vpa1362* and the structural gene *vpa1354* using primers listed in Table 2.2. All PCR screens were performed using strains grown three independent times. The sequence diversity of the T3SS1 and T3SS2 effectors examined was determined by PCR amplifying with the primers listed in Table 2.2 and NEB Phusion HF polymerase with GC buffer as described for the MLST analysis.

Table 2.2. Primers used in this study

Gene	Primer	Amplicon	Sequence	Source
Primers for PCR screen				
tl	tl-F	450 bp	AAAGCGGATTATGCAGAAGCACTG	(6)
	tl-R		GCTACTTTCTAGCATTTTCTCTGC	
tdh	tdh-F	269 bp	GTAAAGGTCTCTGACTTTTGGAC	(6)
	tdh-R		TGGAATAGAACCTTCATCTTCACC	
trh	trh-F	359 bp	CTGAATCMCCAGTTAASGC	(46)
	trh-R		ATGYCCATTKCCGCTCTC	
ORF8	O3MM824	369 bp	AGGACGCAGTTACGCTTGATG	(47)
	O3MM1192		CTAACGCATTGTCCCTTTGTAG	
vp1670	VP1670-F	338 bp	GCGTCACTGCATTCGACAGG	(46)
	VP1670-R		GACTCGTGTGACTCTGCTG	
vp1680	vp1680-F	1-kb	CAGGCATCAGCCAATCCCAATCTTC	This study
	vp1680-R		TCAGTCCGTCAGGCTACCGAG	"
vp1686	vp1686-F	900 bp	GCCATCGGTAAAGCCGTGATAACC	This study
	vp1686-R		TGAAGGCAAACTCAGCATTGGTGG	"
vpa1346	YopP-F	750 bp	CGTCCAACTCTATTGTTGTGATATGGCG	(71)
	vpa1346-R		GGGCTCTGATCTTCGTGAACAG	This study
vpa1354	VPA1354-F	522 bp	CTGGCGAGCCTTCCGTCTC	(46)
	VPA1354-R		CTGGTGAATGGTTCTCCGCAG	
vpa1362	vpa1362-F	800 bp	GAGCAACCGAACGCATCGCT	This study
	vpa1362-R		CGCAGAGAACCAAGAATCTCTGTGG	"
Primers for Sequencing			TGAAGGCAAACTCAGCATTGGTGG	
vp1680	vp1680 M13F	1-kb	TGTAAAACGACGGCCAGTCAGGCATCAGCCAATCCCAATCTTC	This study
	vp1680 M13R		CAGGAAACAGCTATGACCGTCAGTCCGTCAGGCTACCGAG	"
vp1686	vp1686 M13F	900 bp	TGTAAAACGACGGCCAGTGCCATCGGTAAAGCCGTGATAACC	"
	vp1686 M13R		CAGGAAACAGCTATGACCGTGAAGGCAAACTCAGCATTGGTGG	"
vpa1346	vpa1346 M13F	750 bp	TGTAAAACGACGGCCAGTCGTCCAACTCTATTGTTGTGATATGGCG	"
	vpa1346 M13R		CAGGAAACAGCTATGACCGGGGCTCTGATCTTCGTGAACAG	"
vpa1362	vpa1362 M13F	800 bp	TGTAAAACGACGGCCAGTGAGCAACCGAACGCATCGCT	"
	vpa1362 M13R		CAGGAAACAGCTATGACCGCGCAGAGAACCAAGAATCTCTGTGG	"

MLST. Multilocus sequence typing was performed with four housekeeping genes (*recA, gyrB, pyrC, dtdS*) on the *V. parahaemolyticus* environmental (n= 57) and clinical (n= 28) strains using primers previously developed (26). The PCR amplification was performed using NEB Phusion high-fidelity polymerase and products were purified by separation on a 0.7% Seakem LE agarose gel. The products were purified from the gel using a Sigma GenElute gel extraction kit (Sigma Aldrich; St. Louis MO). Sequencing was performed with M13 primers at the Georgia Tech Genome Center on an ABI 3130 Genetic Analyzer. Sequences were assembled in BioEdit (v. 7.0.4.1) (27) and aligned in Mega (v. 4.0) (40).

Sequence and phylogenetic analysis. The nucleotide diversity (π), selection (dN/dS), and number of recombination events of the housekeeping genes and T3SS effectors was determined as previously described (30). The nucleotide diversity (π) of synonymous and nonsynonymous sites and Watterson's theta (θ_ω) were determined using DnaSP (61) as previously described (30). In addition, the number of recombination events within each housekeeping gene was analyzed using the four-gametic test in DnaSP (61). The ratio of nonsynonymous ($\pi_{nonsynon}$) synonymous site changes (π_{synon}), K_a/K_s, was determined as previously described (30). A $K_a/K_s < 1$ indicates strong purifying selection is present.

Serotyping. Serotyping of clinical and environmental strains was performed using O (lipopolysaccharide) and K (capsule) Seiken antiserum (Denka Seiken; Tokyo, Japan) as previously described (18, 22).

PFGE. Pulsed-field gel electrophoresis of the *V. parahaemolyticus* clinical and environmental strains was performed according to the *V. parahaemolyticus* PulseNet protocol (58). Restriction endonuclease profiles were generated using the enzymes SfiI

and NotI. Restricted plugs were run on a BioRad chef mapper using the standard PulseNet settings (36, 58). PFGE patterns were analyzed with Bionumerics (v. 5.0) and fragment sizes were designated relative to the XbaI restricted control strain *Salmonella branderup* H1981.

Plasmid isolation and characterization. A modified Kieser method was used to extract extrachromosomal DNA from *V. parahaemolyticus* clinical and environmental strains as previously described (29). A total of 422 presumptive *V. parahaemolyticus* environmental strains that formed green colonies on TCBS plates were screened for MGE content. In addition, 63 *V. parahaemolyticus* clinical strains provided by the CDC were screened for MGE content. Strains were screened at least two independent times to confirm the presence of MGEs. The Kieser-extracted DNA was visualized by running 80 µl on a 0.7% agarose gel for fours hours at 100 V. Any bands detected by the Kieser method that were smaller or larger than the sheared linear chromosome present at approximately 21-kb were considered MGEs. The *V. parahaemolyticus* clinical and environmental strains selected for MLST analysis were randomly chosen from the strains identified with MGEs and those identified without MGEs to give an approximately even number of strains for comparison.

Swarming motility. The swarming motility was examined for all strains listed in Table 2.1 using methods previously described (75). Briefly, the ATCC strain 17802 that is a non-pandemic O3:K6 strain isolated prior to 1995 was used as a swarm-minus control strain that the swarming of all other strains were measured relative to. The swarming motility was determined by patching strains on heart infusion (HI) swarm plus agar that contained 25g HI, 20g NaCl, and 15g Difco agar per liter (38). Strains were incubated at

30˚C for 48 hours and the area of growth was measured using ImageJ (http://rsb.info.nih.gov/ij/) calibrated to measure in centimeters (cm). The area was determined for each strain three independent times and always compared to a concurrently grown ATCC 17802 strain. The diameter of swarming was determined for each strain and represented as a percentage of the swarm diameter of the swarm$^-$ control strain ATCC 17802. The average of the percentage of 17802 for three independent trials was determined along with the standard deviation, and strains that exhibited swarming ≥ 200% than that of 17802 were considered swarm$^+$ strains (Table 2.1).

TEM analysis. The generation of flagella during growth on plates that promote swarming motility was examined for *V. parahaemolyticus* clinical compared to environmental strains by negative staining and observation by transmission electron microscopy (TEM) as previously performed (56, 65). The bacterial strains were patched into the center of a swarm plates and incubated overnight at 30˚C. Cells were resuspended in phosphate buffered saline and were negatively stained by mixing with 2% phosphotungstic acid (pH 7) and allowed to air dry on a copper-coated grid. The samples were visualized using a JEOL JEM-1210 transmission electron microscope (JEOL, Tokyo) with an accelerating voltage of 80 kV.

Accession numbers. All *recA, gyrB, pyrC, dtdS, vp1680, vp1686, vpa1346,* and *vpa1362* sequences generated in this study are deposited in GenBank under the accession numbers FJ847518-FJ847829.

Results

Population structure of *V. parahaemolyticus* clinical and environmental strains. In order to determine the genetic diversity and population structure of geographically co-occurring and separated *V. parahaemolyticus* environmental strains, we used MLST and PFGE and compared the diversity of the environmental strains to clinical strains isolated from disease-associated samples. In the current study, we define a population as a group of geographically co-occurring strains that would be more likely to exchange genetic material than more geographically separated strains. The strains isolated from each sample site (GA, NC, FL) will be referred to in this study as a population; however, we will likewise address the global population diversity of the *V. parahaemolyticus* species. The *V. parahaemolyticus* environmental strains examined in this study were isolated from sediment and water from the Atlantic Ocean (GA and NC) and the Gulf of Mexico (FL) (Table 2.1). In addition, we isolated strains from the GA sample site in September 2006 and again in September 2007 that were analyzed for changes in the genetic diversity of a localized *V. parahaemolyticus* environmental population over time.

In order to determine the diversity of the *V. parahaemolyticus* environmental strains, we performed MLST with four housekeeping genes. We examined four genes that have previously been used in MLST studies of *V. parahaemolyticus* (26). The four genes were *recA* and *gyrB* from chromosome I and *dtdS* and *pyrC* from chromosome II (Table 2.3). In addition, we examined the nucleotide diversity of T3SS1 and T3SS2 effector protein-encoding genes from select strains. The T3SS1 effector protein-encoding genes analyzed were *vp1680* and *vp1686*, which have been shown to cause

cytotoxicity of eukaryotic cells by inducing autophagy (11, 57). The T3SS2 effector protein-encoding genes analyzed are *vpa1346* and *vpa1362*.

To determine whether the level of sequence diversity in each of the four housekeeping genes and T3SS effectors examined varied among the clinical and environmental strains, we analyzed the nucleotide diversity (π) independently for clinical compared to environmental strains. The nucleotide diversity (π) measures the average number of nucleotide differences per site when comparing sequences from strains within a population (50). The nucleotide diversity was analyzed separately for synonymous and nonsynonymous sites to determine the contribution to diversity of each of these sites. Watterson's theta (θ_ω) is a measure of the mutation rate per nucleotide within a population (73). The nucleotide diversity (π) of the housekeeping genes analyzed was similar for *V. parahaemolyticus* environmental strains compared to clinical strains (Table 2.3). The nucleotide diversity at synonymous and nonsynonymous sites in *recA* and *dtdS* alleles of both clinical and environmental strains was more than double that of the *gyrB* and *pyrC* alleles examined (Table 2.3). In addition, the nucleotide diversity (π) at synonymous and nonsynonymous sites was higher for the T3SS1 effector protein-encoding gene *vp1680* than for the effector protein-encoding gene *vp1686* (Table 2.3). The nucleotide diversity of the T3SS2 effector protein-encoding genes *vpa1346* and *vpa1362* was similar and significantly lower than other genes analyzed (Table 2.3).

Table 2.3. Nucleotide diversity of housekeeping genes and T3SS effector protein-encoding genes.

						p				
Gene	Size (bp)	No. sequences	No. alleles	No. polymorphic sites	q_w	Synonymous	Non-synonymous	Total	K_a/K_s	Recombination
recA (Chr I)	642	85	44	67	0.01881 (0.02027)	0.10161 (0.07931)	0.00043 (0.00097)	0.02477 (0.01982)	0.0042 (0.0122)	8
gyrB (Chr I)	588	85	47	51	0.01311 (0.01623)	0.04784 (0.04439)	0.00047 (0.00039)	0.01132 (0.01047)	0.0098 (0.0087)	9
pyrC (Chr II)	531	85	47	47	0.01258 (0.01756)	0.03857 (0.04566)	0.00198 (0.00141)	0.01076 (0.01203)	0.0513 (0.0308)	7
dtdS (Chr II)	489	85	46	46	0.01997 (0.01729)	0.09073 (0.09771)	0.00058 (0.00009)	0.02233 (0.02367)	0.0063 (0.0009)	12
vp1680 (Chr I)	696	23	16	696	0.01815 (0.02617)	0.06873 (0.08014)	0.00172 (0.00336)	0.01792 (0.022197)	0.025 (0.0419)	13
vp1686 (Chr I)	705	23	17	42	0.01597 (0.01308)	0.03685 (0.04026)	0.00661 (0.00509)	0.01354 (0.01315)	0.1794 (0.1264)	8
vpa1346 (Chr II)	744	6	4	28	(0.01648)	(0.03137)	(0.00775)	(0.01281)	(0.247)	0
vpa1362 (Chr II)	720	6	4	12	(0.00730)	(0.01387)	(0.00338)	(0.00583)	(0.2436)	0

[a]Values listed in parentheses represent clinical strains that were isolated from disease-associated samples.

In addition, we measured the selective pressure on each gene analyzed by measuring the level of change in the nonsynonymous sites to change in synonymous sites (dN/dS). The ratio of nonsynonymous changes to synonymous changes (K_a/K_s) for each gene analyzed were much less than one indicating these genes are all subject to purifying selection (Table 2.3). The number of recombination events within each housekeeping gene was detected using the four-gametic test in DnaSP (61). There were from 7-12 recombination events detected in the four housekeeping genes analyzed (Table 2.3). There were no recombination events detected in the T3SS2 genes analyzed (Table 2.3).

Phylogenetic analysis of a concatenation of four housekeeping genes revealed that the *V. parahaemolyticus* strains examined formed a monophyletic group that was distinct from *V. harveyi* (Figure 2.1). There were two separate sub-groups present within the *V. parahaemolyticus* group (Figure 2.1). The first group was composed primarily of clinical strains while the majority of the strains in the other group were of environmental origin (Figure 2.1). The high genetic diversity of each gene reflected by the significant number of sequence types (Table 2.3) was evident in the concatenated phylogeny (Figure 2.1). There were four separate groups present in the tree that were composed of three or more strains that were closely related as indicated by a bootstrap value ≥99 (Figure 2.1 A-D). Of the four groups of related strains, the first was composed of the O3:K6 clonal pandemic strains and one environmental strain, AF91, that was isolated from sediment in FL and had the O3:K6 serotype (Figure 2.1 A). The second group of strains consisted only of O4:K12 clinical strains (Figure 2.1 B), while the third group was a mixture of environmental strains isolated from NC and GA (Figure 2.1 C) and the fourth group was composed entirely of strains isolated from FL (Figure 2.1 D). In addition, there were five

doublets of strains that had high similarity (bootstrap ≥99) (Figure 2.1 E-I). The first two doublets of related strains were clinical strains that had similar serotypes (Figure 2.1 E, F). The third doublet was formed by two environmental strains isolated from NC (Figure 2.1 G). The fourth doublet consisted of a clinical strain and an environmental strain isolated from GA (Figure 2.1 H). The last doublet identified consisted of two strains isolated from FL (Figure 2.1 I).

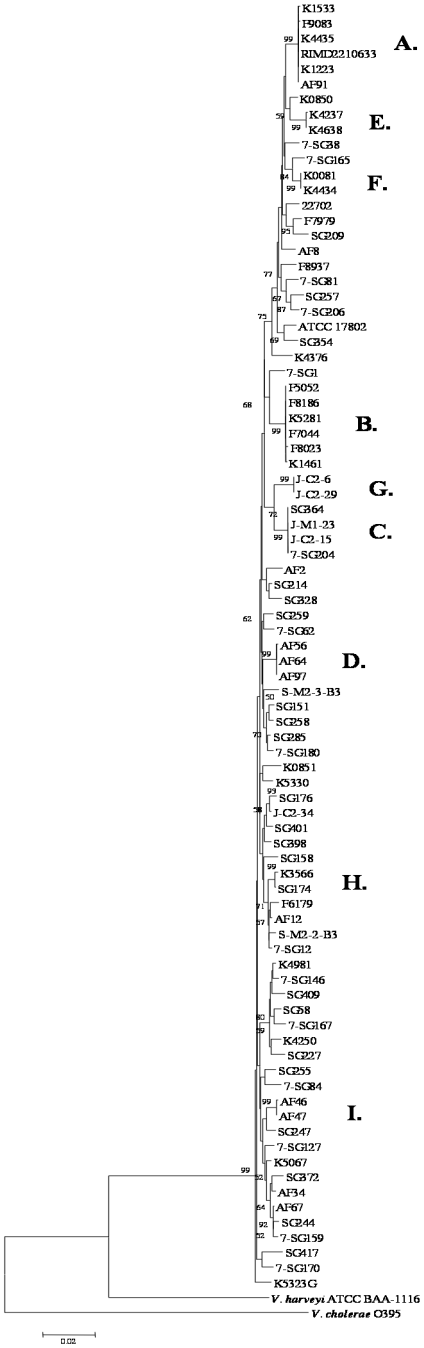


Figure 2.1. A neighbor-joining tree constructed from a concatenation of four housekeeping genes (*recA, gyrB, pyrC, dtdS*) from *V. parahaemolyticus* clinical and environmental strains. The phylogeny was constructed using the Kimura 2-parameter model with 1,000 bootstrap replications. The scale represents the evolutionary distance of 0.02 nucleotide changes. Labels A-D indicate four groups of three or more strains that have high similarity. Labels E-I indicate five doublets of strains that are closely related.

Phylogenetic analysis of the individual gene trees revealed the level of sequence divergence varied by gene (Figures 2.2-2.5). In each of the individual phylogenies the *V. parahaemolyticus* strains formed a monophyletic group that was distinct from *V. harveyi* (Figures 2.2-2.5). The *recA* (Figure 2.2) and *dtdS* (Figure 2.5) phylogenies had deeper branching indicating greater sequence divergence among the strains, while the *gyrB* (Figure 2.3) and *pyrC* (Figure 2.4) phylogenies showed these genes were conserved among *V. parahaemolyticus* clinical and environmental strains. The *dtdS* phylogeny revealed a different topology for the O4:K12 clonal strains that was mixed with some environmental strains (Figure 2.5) rather than being composed only of the O4:K12 strains as in the *recA* (Figure 2.2), *gyrB* (Figure 2.3), and *pyrC* (Figure 2.4) phylogenies. Similarly, the O3:K6 clonal strains were mixed with environmental strains in the *gyrB* phylogeny whereas they formed a group composed only of O3:K6 strains in the *recA* (Figure 2.2), *pyrC* (Figure 2.4), and *dtdS* (Figure 2.5) phylogenies. The O3:K6 group in each of the individual gene trees included the *V. parahaemolyticus* environmental strain AF91 that was isolated from a FL sediment sample (Figures 2.2-2.5). In addition to AF91 in the *gyrB* phylogeny, the O3:K6 group included the environmental strains *V. parahaemolyticus* SG151, SG258, and 7-SG1 (Figure 2.3), which were isolated from GA sediment in 2006 and 2007, respectively.

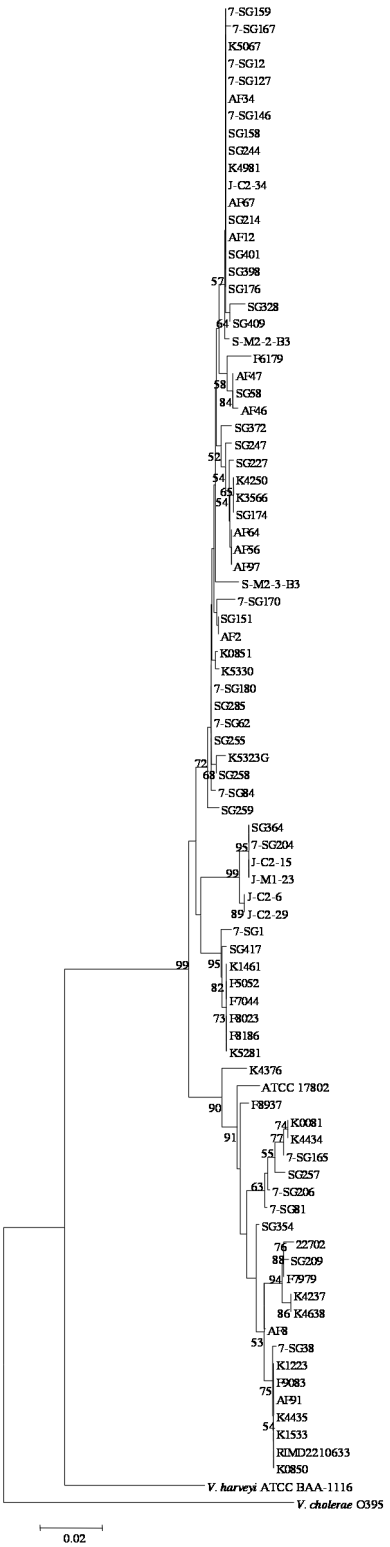


Figure 2.2. The neighbor-joining phylogeny of *recA* nucleotide sequences from *V. parahaemolyticus* clinical and environmental strains examined in this study. The scale represents the evolutionary distance of 0.02 nucleotide changes.

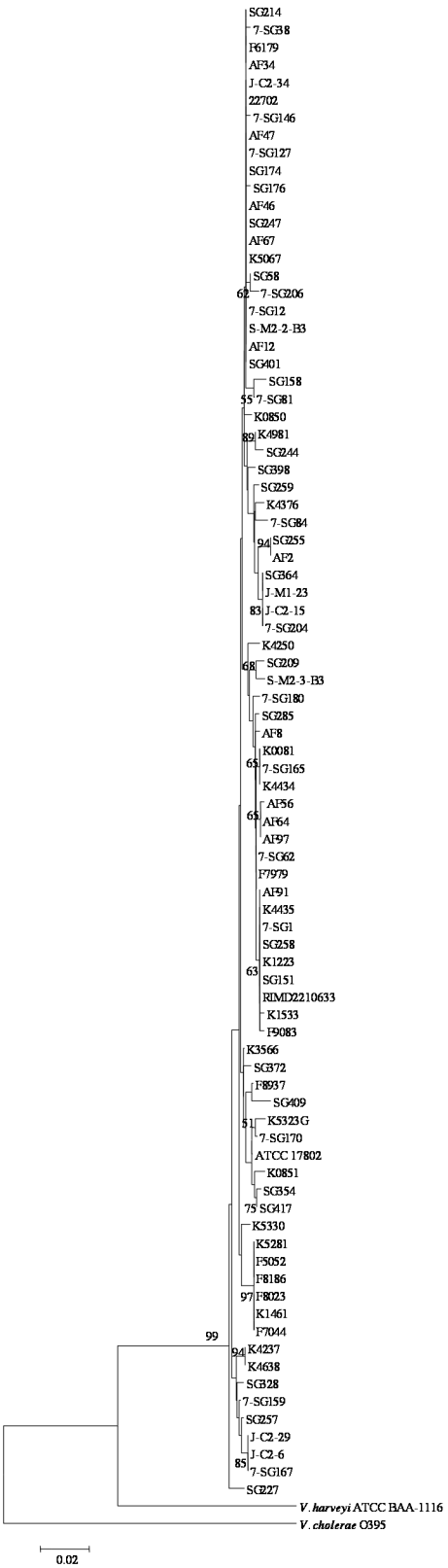


Figure 2.3. The neighbor-joining phylogeny of *gyrB* nucleotide sequences from *V. parahaemolyticus* clinical and environmental strains examined in this study. The scale represents the evolutionary distance of 0.02 nucleotide changes.

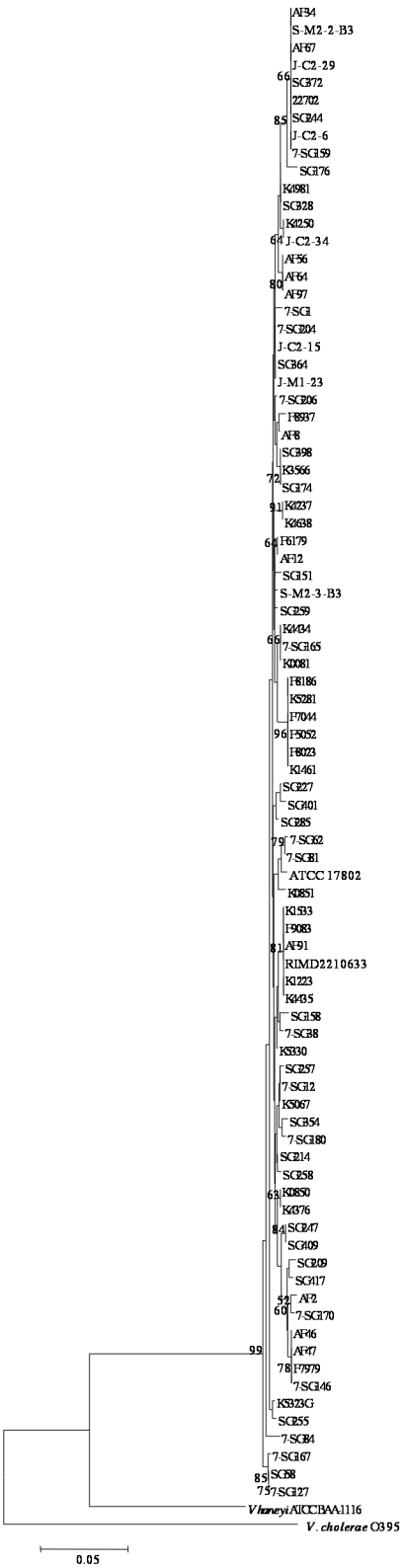


Figure 2.4. The neighbor-joining phylogeny of *pyrC* nucleotide sequences from *V. parahaemolyticus* clinical and environmental strains examined in this study. The scale represents the evolutionary distance of 0.05 nucleotide changes.

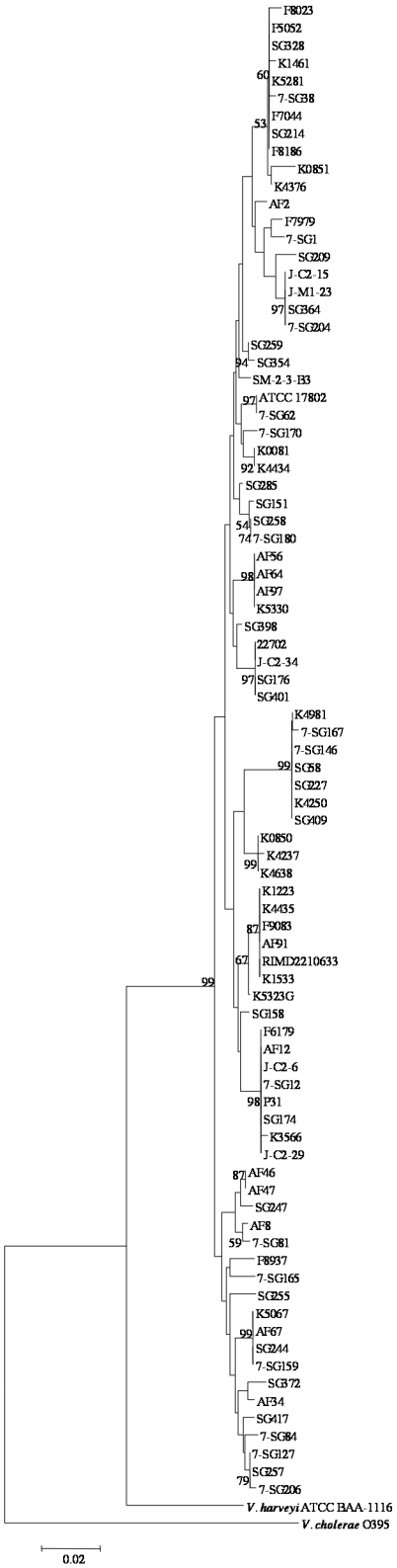


Figure 2.5. The neighbor-joining phylogeny of *dtds* nucleotide sequences from *V. parahaemolyticus* clinical and environmental strains examined in this study. The scale represents the evolutionary distance of 0.02 nucleotide changes.

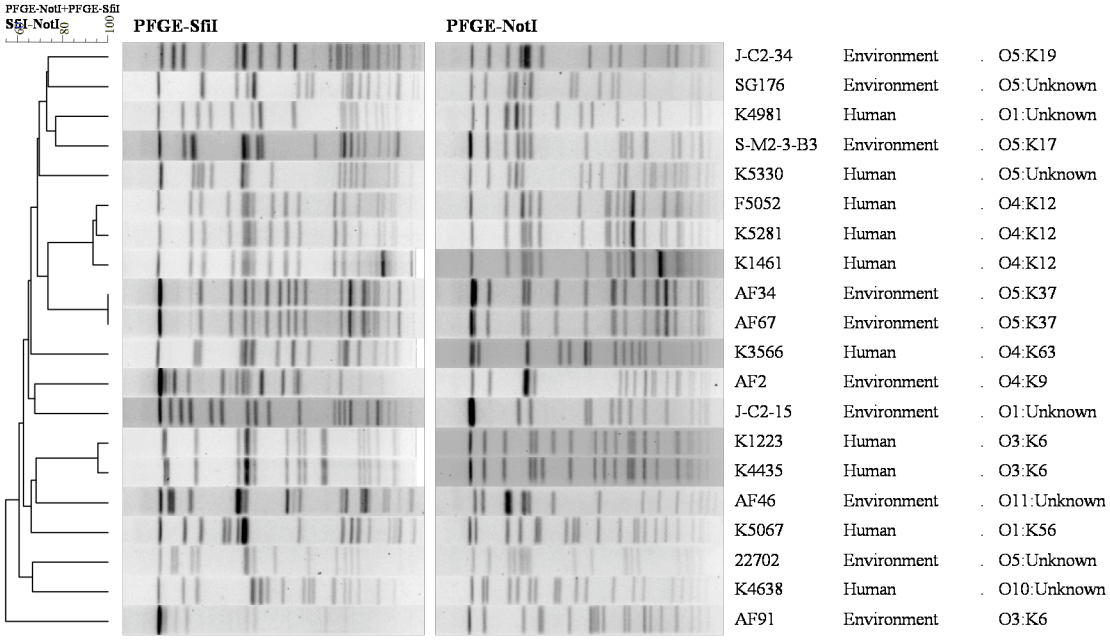


Figure 2.6. Dendrogram and PFGE patterns of NotI and SfiI-digested DNA from *V. parahaemolyticus* clinical and environmental strains examined in this study.

There were unique PFGE patterns for almost all of the 10 *V. parahaemolyticus* environmental strains examined. The environmental strains AF34 and AF67, both isolated from FL, exhibited identical PFGE patterns. Of the 10 *V. parahaemolyticus* clinical strains examined, the three O4:K12 strains and two O3:K6 pandemic strains exhibited similar profiles as indicated by the dendrogram. The *V. parahaemolyticus* environmental strain AF91 was serotyped as an O3:K6 strain; however, the PFGE pattern of this strain was not similar to the clonal pandemic O3:K6 strains examined. The AF91 strain was part of a larger group within the tree that included the O3:K6 pandemic strains, two additional environmental strains, and two non-clonal clinical strains.

Distribution of virulence-associated genes among *V. parahaemolyticus* strains. The presence of all previously identified virulence-associated genes was examined by PCR analysis. As a confirmation of whether strains are *V. parahaemolyticus* we screened for the presence of the thermolabile hemolysin (*tl*), which has been shown to occur in all *V. parahaemolyticus* strains (6, 46); however, the contribution of this hemolysin to virulence is not known. To determine the presence of all demonstrated virulence-associated genes we screened for the thermostable direct hemolysins *tdh* and *trh* and both of the type III secretion systems (T3SS1 and T3SS2). In addition, we examined the frequency of the putative virulence factor, ORF8, encoded by the filamentous phage f237.

The *tl* gene was detected among all the *V. parahaemolyticus* clinical and environmental strains examined (Table 2.1). Of the known virulence-associated genes examined, all were detected more frequently among the clinical strains than the environmental strains with the exception of T3SS1, which was present in all strains examined (Tables 2.1, 2.4). The T3SS2 genes were detected among all of the O3:K6 clinical strains and none of the environmental strains examined (Tables 2.1, 2.4). In addition, there was an O4:K10 strain F8937 that had the T3SS2 genes but lacked *tdh* and *trh* (Table 2.1). *V. parahaemolyticus* F8937 was identified within the large group of strains that included the pre-1995 O3:K6 clinical strain ATCC 17802 and several environmental strains in addition to the sub-group of post-1995 O3:K6 pandemic strains (Figure 2.1). The pandemic phage gene ORF8 was present in all of the O3:K6 pandemic strains analyzed; however, this gene was also detected in 13% of the non-O3:K6 clinical strains (Table 2.4). All of the O4:K12 clonal strains examined had both *tdh* and *trh* but lacked the T3SS2 genes (Table 2.1) and represented in Table 2.4 as the clonal non-O3:K6

strains. The *tdh* and *trh* genes were detected among 0% and 3% of the *V. parahaemolyticus* environmental strains, respectively (Table 2.4).

Table 2.4. Virulence-associated genes of *V. parahaemolyticus* clinical and environmental strains

		No. Strains with Virulence Genes (%)					
		tl	ORF8	*tdh*	*trh*	T3SS1	T3SS2
Clinical	Clonal O3:K6	4	4 (100)	4 (100)	0 (0)	4 (100)	4 (100)
	Clonal Non-O3:K6	6	1 (17)	6 (100)	6 (100)	6 (100)	0 (0)
	Non-clonal	17	2 (12)	6 (35)	5 (29)	17 (100)	1 (6)
	Total Clinical	27	7 (26)	16 (59)	11 (41)	27 (100)	5 (19)
Environmental	Total	59	0 (0)	0 (0)	2 (3)	59 (100)	0 (0)

Phenotypic diversity of *V. parahaemolyticus* strains. In order to determine the potential for *V. parahaemolyticus* pathogen emergence from environmental populations, we examined the distribution of virulence-associated genes, MGEs, and serotype diversity (Table 2.2). Of 422 presumptive *V. parahaemolyticus* environmental strains and 63 *V. parahaemolyticus* clinical strains that were screened for the presence of MGEs, there were 157 environmental strains (37%) and 18 clinical strains (29%) that had detectable bands (Table 2.5). Among the 85 *V. parahaemolyticus* strains selected for MLST analysis there were 32 (56%) environmental strains and 9 (39%) clinical strains with detectable MGEs.

Table 2.5. Summary of MGEs detected among *V. parahaemolyticus* clinical and environmental strains

	No. Strains with Extrachromosomal Bands (%)				
	n	≤23-kb	23-60-kb	≥60-kb	Total
Oysters	11	1 (9)	1 (9)	0	-
Sediment	166	32 (19)	11 (7)	30 (18)	-
Water	245	44 (18)	9 (4)	29 (12)	-
Total Environmental	422	77 (18)	21 (5)	59 (14)	157 (37)
Clinical	63	9 (14)	6 (10)	3 (5)	18 (29)
Totals	485	86 (18)	27 (6)	62 (13)	175 (36)

In addition, we determined the frequency of swarming motility among *V. parahaemolyticus* clinical and environmental strains (Table 2.6). A previous study demonstrated that the O3:K6 pandemic strains exhibited significant swarming motility (75). We confirm previous findings that swarming motility was more frequent among the O3:K6 clinical strains (90%) than among the non-O3:K6 clinical strains (18%)(Table 2.6; p= 0.00003). In addition, we show that swarming motility is highly variable among environmental strains based on isolation site (0-86%). Six of the seven strains examined from NC exhibited significant swarming motility (86%), while none of the 15 strains isolated from GA in 2007 exhibited swarming motility (Table 2.6).

Table 2.6. Swarming motility of *V. parahaemolyticus* clinical and environmental strains

Source	Type	n	No. Strains (%)
Clinical	Total	32	14 (44)
	O3:K6	11	10 (90)
	Non-O3:K6	22	4 (18)
Environmental	Total	60	18 (30)
	GA 2006	26	10 (38)
	GA 2007	15	0
	FL	11	1 (9)
	NC	7	6 (86)

To confirm whether the swarming motility of *V. parahaemolyticus* pandemic strains is associated with the production of multiple flagella we used transmission electron microscopy (TEM) to analyze cells that had been growing on swarm agar (Figure 2.7). In addition, we looked for the presence of multiple flagella among an environmental strain that swarms and the non-swarming environmental strain AF91 that resembles the *V. parahaemolyticus* pandemic strains (Figure 2.7). We detected cells with multiple flagella for the pandemic strain *V. parahaemolyticus* RIMD2210633 and the environmental strain 22702, which both exhibit swarming motility (Figure 2.7). There were no multi-flagellated cells identified for the non-pandemic strain *V. parahaemolyticus* ATCC 17802 (Figure 2.7). In contrast, the environmental strains *V. parahaemolyticus* 22702 and AF91 both had cells with multiple flagella (Figure 2.7).

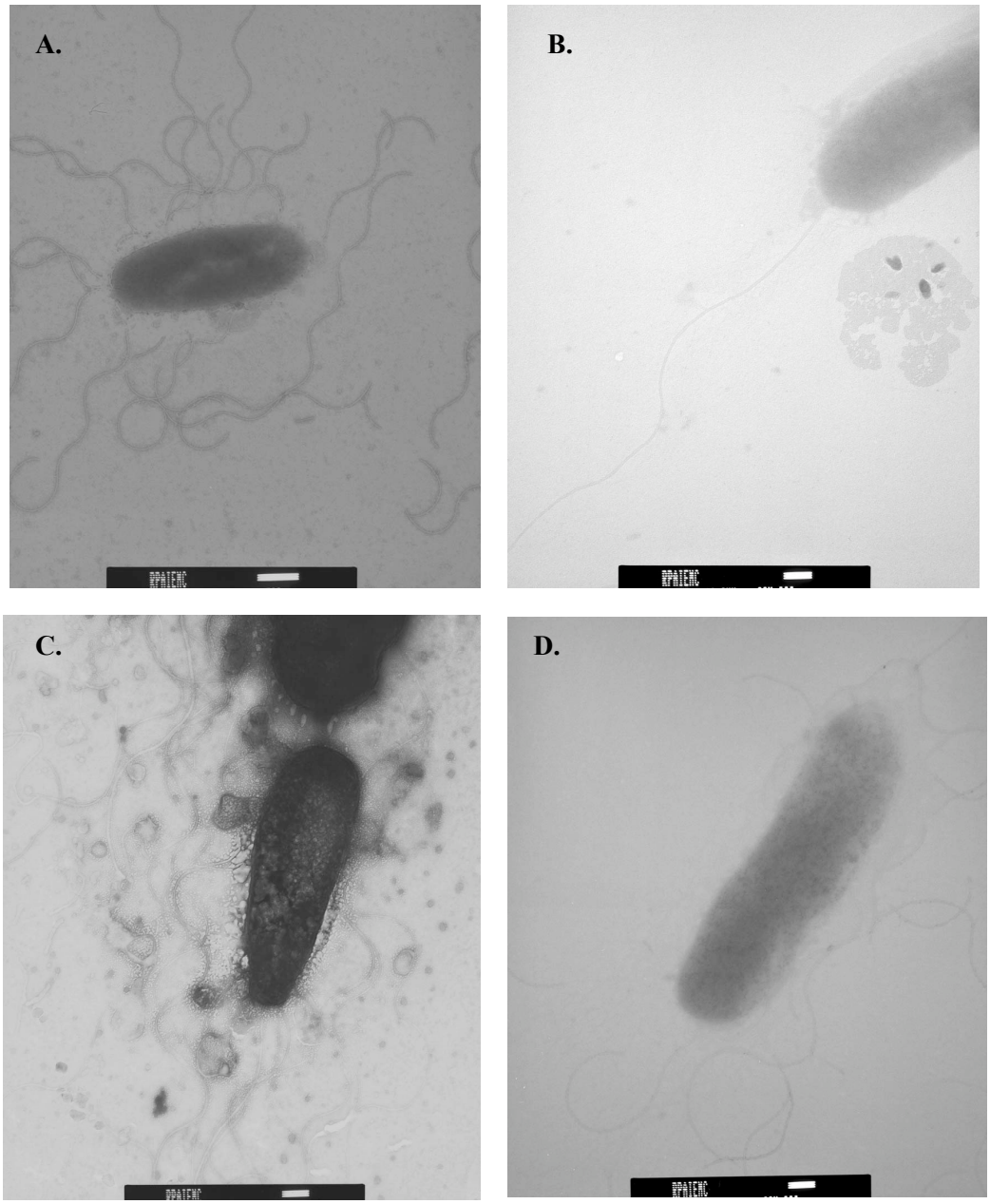
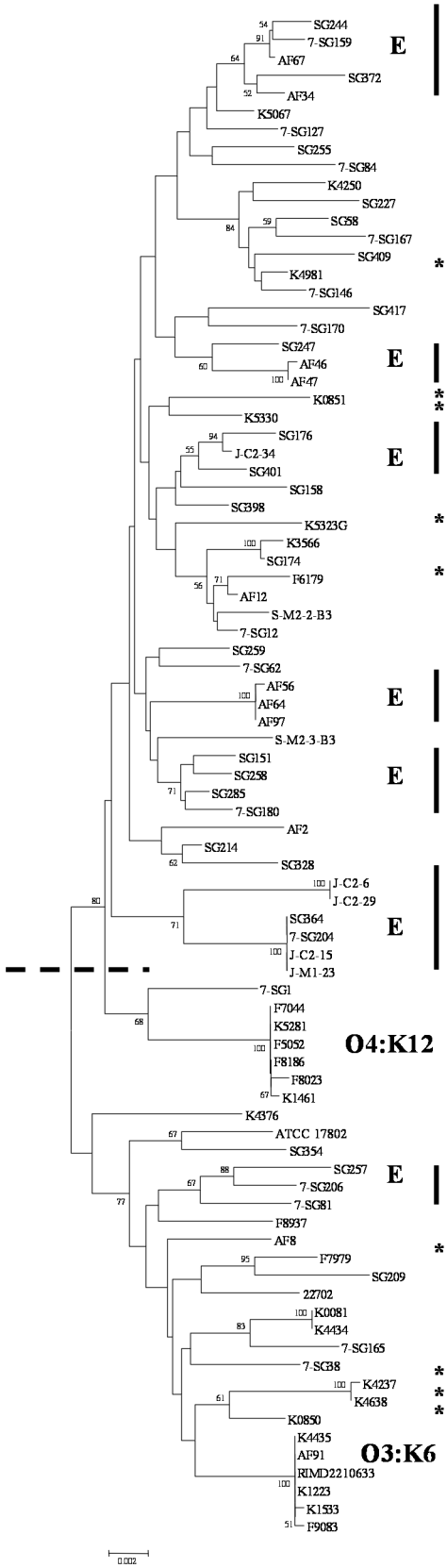


Figure 2.7. Transmission electron micrographs of *V. parahaemolyticus* strains grown on swarm agar. Representative cells with single or multiple flagella of the *V. parahaemolyticus* O3:K6 pandemic strain RIMD2210633 (A), the non-pandemic O3:K6 strain ATCC 17802 (B), the non-swarming O3:K6 environmental strain AF91 (C), and the swarming environmental strain 22702. The scale bar represents 500 nm (A), 200 nm (B), nm (C), and nm (D).

Comparison of the concatenated MLST phylogeny with virulence-associated genes, MGE content, serotypes, isolation source, and swarming motility showed there were groups of strains formed that shared similar traits (Figure 2.8). The *V. harveyi* and *V. cholerae* sequences included in Figure 2.1 were excluded from this tree to better show the relationship among the *V. parahaemolyticus* strains. Overall, the tree topology remains the same as that present in Figure 2.1. Analysis of the isolation source relative to the tree topology indicated that the tree can roughly be divided into two large groups as discussed for Figure 2.1. The first large group consists of 20 of the 28 total clinical strains examined and only 11 of the 57 environmental strains examined (Figure 2.8). In addition, there were seven strongly associated sub-groups of environmental strains present in the phylogeny, labeled with an "E" (Figure 2.8). Of these environmental sub-groups there were three that were composed of strains from multiple sample sites, while the remaining four groups were composed of strains isolated only from GA or only from FL (Figure 2.8). Some of these related strains isolated from the same sample site had differences in motility or MGE content indicating they are likely unique strains or have recently diverged from a common ancestor. There were several groups of strains that were capable of swarming motility including the O3:K6 clonal group and an environmental group of strains isolated from GA and NC (Figure 2.8). In addition, there were groups of strains that had detectable MGEs (Figure 2.8).

The virulence-associated genes were mostly present within the clinical group; however, several clinical strains that possessed one or both of the hemolysins grouped with the environmental strains (Figure 2.8). In addition, there were nine clinical strains that did not have either of the hemolysins or T3SS2 genes (Table 2.1) that are indicated

in Figure 2.8. Several of the clinical strains that lacked hemolysins formed two separate groups (Figure 2.8). In particular, there was a group of three clinical strains lacking virulence genes (K4237, K4638, and K0850) that formed a branch with the pandemic O3:K6 strains (Figure 2.8). Within the O3:K6 group was the *V. parahaemolyticus* environmental strain AF91 as discussed in Figure 2.1. Of the strains within the O3:K6 group, AF91 lacked the virulence-associated genes and swarming motility that was detected for the other strains (Figure 2.8). The *V. parahaemolyticus* O3:K6 strain K4435 was the only strain that did not have swarming motility; however, this strain exhibited variable swarming and can be considered a weak swarmer.

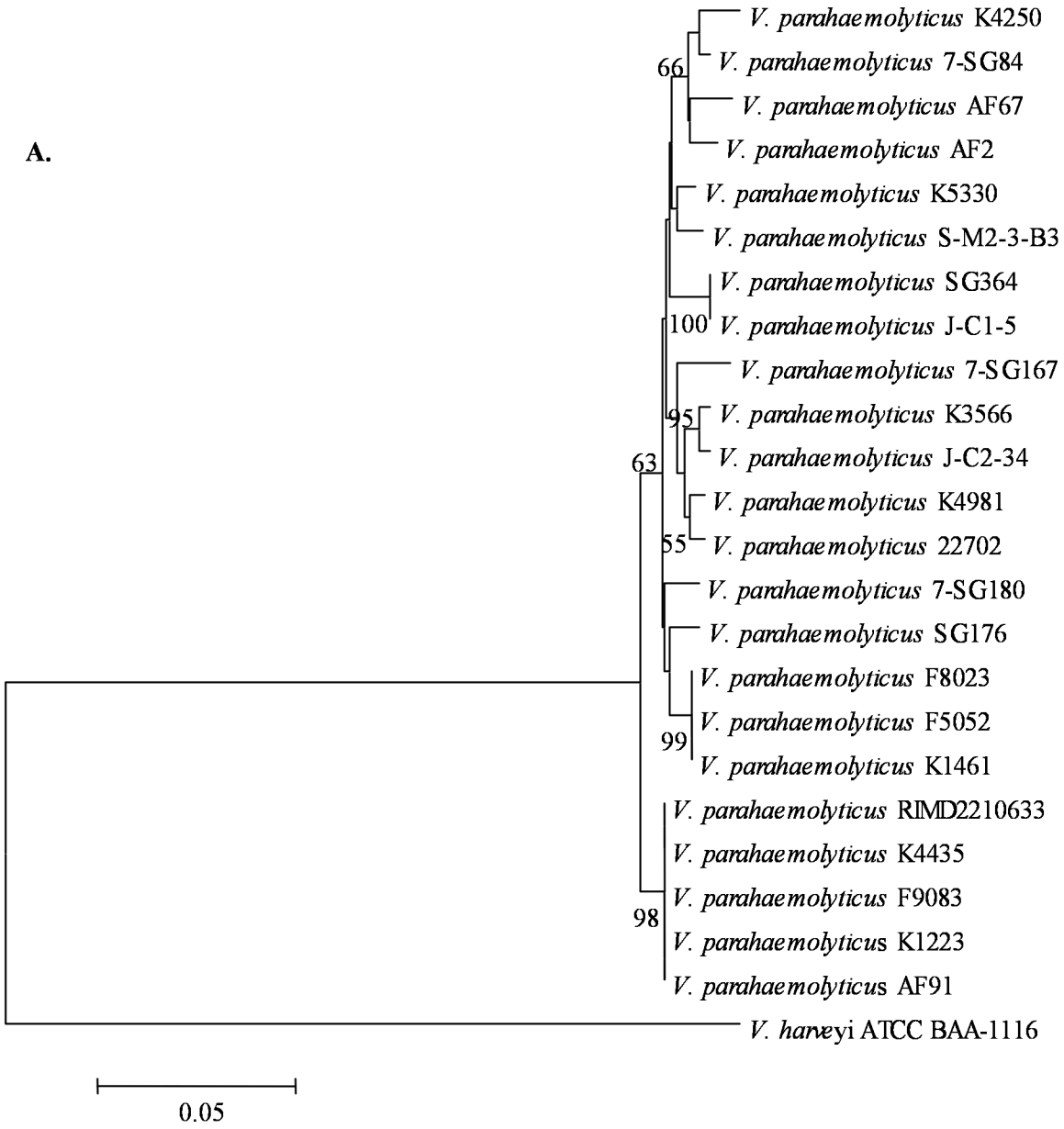


	Strain Id	Serotype	Source	MGEs	ORF8	*tdh*	*trh*	T3SS1	T3SS2	Swarming
	SG244	O6:Kuk	W	+	-	-	-	+	-	+
	7-SG159	ND	S	-	-	-	-	+	-	-
	AF67	O5:K37	S	+	-	-	-	+	-	-
	SG372	O5:Kuk	S	+	-	-	-	+	-	+
	AF34	O5:K37	S	+	-	-	-	+	-	-
	K5067	O1:K56	C	+	-	+	+	+	-	-
	7-SG127	ND	S	-	-	-	-	+	-	-
	SG255	ND	S	-	-	-	-	+	-	-
	7-SG84	ND	S	-	-	-	-	+	-	-
	K4250	O4:K63	C	+	-	+	+	+	-	-
	SG227	ND	S	-	-	-	-	+	-	+
	SG58	ND	W	-	-	-	-	+	-	-
	7-SG167	ND	S	+	-	-	-	+	-	-
	SG409	O1:K33	S	+	-	-	-	+	-	+
*	K4981	O1:Kuk	C	-	+	-	-	+	-	-
	7-SG146	ND	S	-	-	-	-	+	-	-
	SG417	O4:K67	S	+	-	-	-	+	-	-
	7-SG170	ND	S	-	-	-	-	+	-	-
	SG247	ND	W	-	-	-	-	+	-	-
	AF46	O11:Kuk	W	+	-	-	-	+	-	-
	AF47	ND	W	+	-	-	-	+	-	+
*	K0851	O1:Kuk	C	-	-	-	-	+	-	+
*	K5330	O5:Kuk	C	-	-	-	-	+	-	-
	SG176	O5:Kuk	W	+	-	-	-	+	-	+
	J-C2-34	O5:K19	S	+	-	-	-	+	-	+
	SG401	O5:K61	S	+	-	-	-	+	-	+
	SG158	ND	S	-	-	-	-	+	-	-
	SG398	ND	S	-	-	-	-	+	-	-
*	K5323 g	O5:K17	C	-	-	-	-	+	-	-
	K3566	O4:K63	C	-	-	+	+	+	-	-
	SG174	ND	W	-	-	-	-	+	-	-
*	F6179	O6:K18	C	+	-	-	-	+	-	-
	AF12	ND	S	-	-	-	-	+	-	-
	S-M2-2-B3	ND	S	-	-	-	-	+	-	+
	7-SG12	ND	S	-	-	-	-	+	-	-
	SG259	ND	S	+	-	-	+	+	-	-
	7-SG62	ND	S	+	-	-	-	+	-	-
	AF56	O6:K37	S	+	-	-	-	+	-	-
	AF64	ND	S	+	-	-	-	+	-	-
	AF97	ND	W	-	-	-	-	+	-	-
	S-M2-3-B3	O5:K17	S	+	-	-	-	+	-	+
	SG151	ND	W	-	-	-	-	+	-	-
	SG258	O1:Kuk	S	+	-	-	-	+	-	+
	SG285	O10:Kuk	S	+	-	-	-	+	-	-
	7-SG180	ND	S	-	-	-	-	+	-	-
	AF2	O4:K9	O	+	-	-	-	+	-	-
	SG214	O1:K9	W	+	-	-	-	+	-	-
	SG328	ND	W	+	-	-	-	+	-	+
	J-C2-6	ND	S	-	-	-	-	+	-	-
	J-C2-29	ND	S	-	-	-	-	+	-	+
	SG364	O1:Kuk	S	+	-	-	-	+	-	+
	7-SG204	ND	S	+	-	-	-	+	-	-
	J-C2-15	O1:Kuk	S	+	-	-	-	+	-	+
	J-M1-23	O1:K32	S	+	-	-	-	+	-	+
	7-SG1	ND	S	-	-	-	-	+	-	-
	F7044	O4:K12	C	-	-	+	+	+	-	-
	K5281	O4:K12	C	-	-	+	+	+	-	-
	F5052	O4:K12	C	-	+	+	+	+	-	-
	F8186	O4:K12	C	-	-	+	+	+	-	-
	F8023	O4:K12	C	-	-	+	+	+	-	+
	K1461	O4:K12	C	+	-	+	+	+	-	-
	K4376	O8:K21	C	+	-	+	+	+	-	+
	17802	O3:K6	C	-	-	+	-	+	-	-
	SG354	O2:K28	W	+	-	-	-	+	-	-
	SG257	O8:Kuk	S	+	-	-	+	+	-	-
	7-SG206	ND	S	-	-	-	-	+	-	-
	7-SG81	ND	S	+	-	-	-	+	-	-
	F8937	O4:K10	C	+	-	-	-	+	+	-
	AF8	ND	S	-	-	-	-	+	-	-
*	F7979	O5:K17	C	+	+	-	-	+	-	-
	SG209	O2:K28	S	+	-	-	-	+	-	-
	22702	O5:Kuk	S	+	-	-	-	+	-	+
	K0081	O6:K18	C	+	-	+	+	+	-	-
	K4434	O6:K18	C	-	-	-	-	+	-	-
	7-SG165	ND	S	+	-	-	-	+	-	-
	7-SG38	ND	S	-	-	-	-	+	-	-
*	K4237	O10:K56	C	+	-	-	-	+	-	+
*	K4638	O10:Kuk	C	-	-	-	-	+	-	-
*	K0850	O1:K25	C	-	-	-	-	+	-	-
	K4435	O3:K6	C	-	+	+	-	+	+	-
	AF91	O3:K6	S	-	-	-	-	+	-	-
	RIMD2210633	O3:K6	C	+	+	+	-	+	+	+
	K1223	O3:K6	C	+	+	+	-	+	+	+
	K1533	O3:K6	C	-	+	+	-	+	+	+
	F9083	O3:K6	C	-	+	+	-	+	+	+

Figure 2.8. A comparison of the concatenated MLST neighbor-joining phylogenetic analysis of four housekeeping genes presented in Figure 2.1 to strain isolation source, serotype, MGE content, virulence-associated gene content, and motility data. The tree was generated using the Kimura 2-parameter model and 1,000 bootstrap replications. The *V. harveyi* and *V. cholerae* outgroups included in the Figure 2.1 phylogeny were not included in this tree. In the descriptions of the sources, "C" represents clinical, "S" represents sediment, "W" represents water, and "O" represents oyster. The shaded boxes represent the following: clinical source, light yellow; sediment, light green; water, light blue; MGE +, gray; ORF8 +, orange; hemolysin +, blue; T3SS2 +, yellow; swarm +, green. An "E" denotes an environmental group and the large primarily clinical group is below the dashed line. The O4:K12 and O3:K6 clonal groups are indicated. An asterisk * indicates clinical strains that do not have either of the hemolysins or T3SS2.

Diversity of T3SS effectors among *V. parahaemolyticus* clinical and environmental strains. As shown in Tables 2.1 and 2.4 the T3SS1 genes were detected in all *V. parahaemolyticus* clinical and environmental strains examined in this study. Phylogenetic analysis of the effectors *vp1680* and *vp1686* from select strains (Figure 2.9) reflected the sequence diversity reported for these genes in Table 2.3. The *V. parahaemolyticus* environmental strain AF91 that grouped with the O3:K6 clonal strains in the concatenated housekeeping gene phylogeny and single gene trees (Figures 2.1-2.5) likewise formed a group with the O3:K6 clonal strains in the *vp1680* and *vp1686* phylogenies (Figure 2.9). All the *V. parahaemolyticus* T3SS1 sequences analyzed formed a monophyletic group relative to the *V. harveyi* T3SS1 orthologs included in the

trees as outgroups (Figure 2.9). The clinical O3:K6 strains and O4:K12 strains formed separate groups distinct from the rest of the strains. Overall, the T3SS1 gene trees (Figure 2.9) had a similar topology relative to the MLST phylogeny (Figure 2.1).



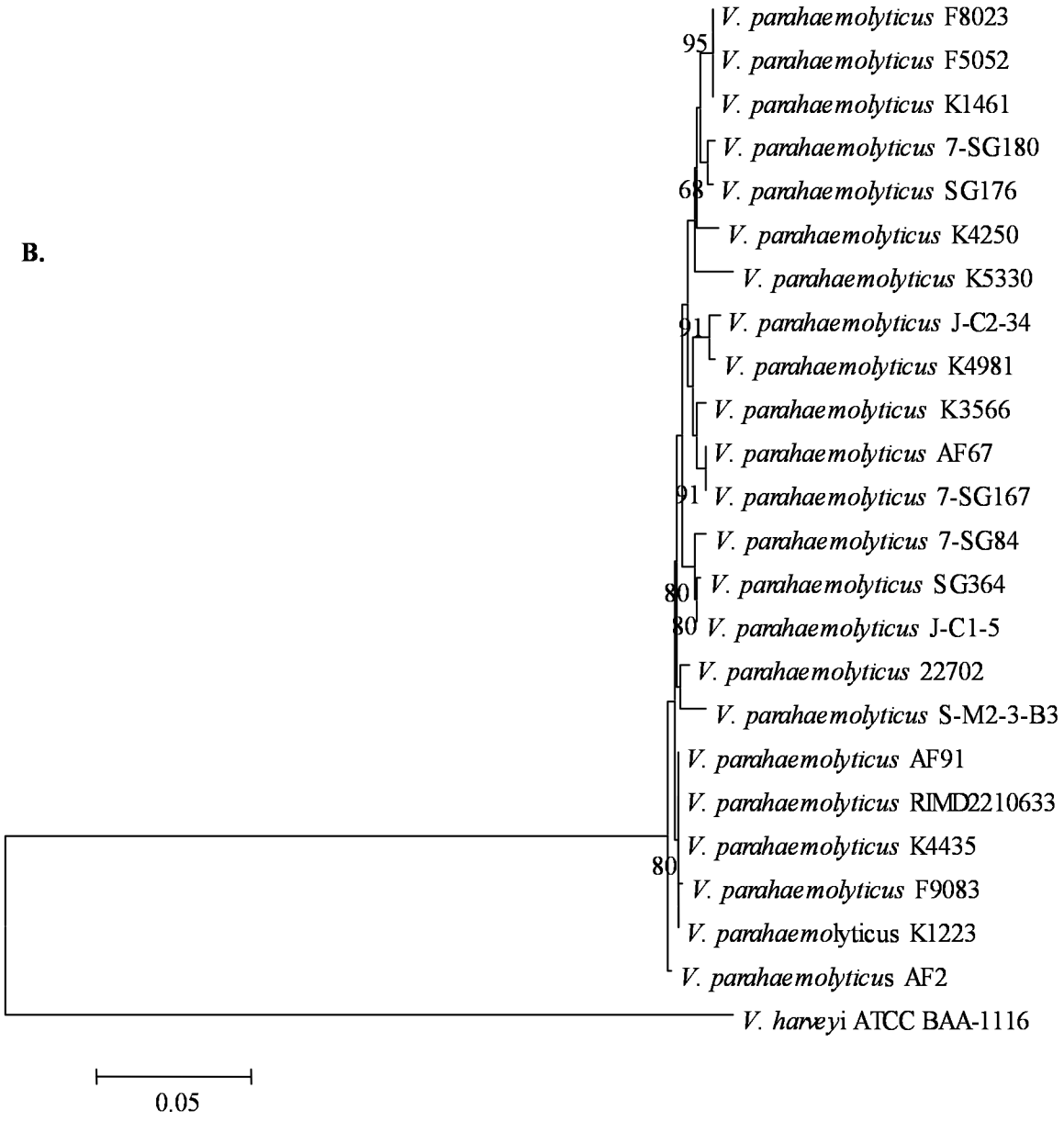


Figure 2.9. Neighbor-joining analysis of the nucleotide sequences of the T3SS1 effector protein-encoding genes *vp1680* (A) and *vp1686* (B) from select *V. parahaemolyticus* clinical and environmental strains. The tree was constructed using the Kimura 2-parameter model and 1,000 bootstrap replications. Only bootstrap values ≥50 are shown. The scale bar represents the evolutionary distance of 0.05 nucleotide changes.

In contrast, sequence analysis of the T3SS2 effectors *vpa1346* and *vpa1362* from the *V. parahaemolyticus* O3:K6 strains one O4:K10 strain revealed these genes were significantly conserved among strains (Table 2.3). Phylogenetic analysis of the T3SS2 effector protein-encoding genes *vpa1346* (Figure 2.10 A) and *vpa1362* (Figure 2.10 B) confirmed the sequence similarity reported in Table 2.3. The T3SS2 effectors analyzed from the O3:K6 strains formed a related group that was separate from the effectors of the O4:K10 strain F8937 (Figure 2.10). An ortholog of the *vpa1346* and *vpa1362* genes was identified in the *V. parahaemolyticus* TH3996 genomic island (T3SS2β) associated with *trh* and was included in the phylogenies for comparison with the T3SS2α genes. In addition, an ortholog of the *vpa1362* effector protein-encoding gene was identified for the *V. cholerae* non-O1/non-O139 strain AM-19226 and was included in the phylogeny; however, no *V. cholerae* genes were identified that had similarity to *vpa1346*. The *vpa1346* T3SS2β ortholog was included in the phyogenetic analysis and formed an outgroup that was distinct from the *V. parahaemolyticus* T3SS2α genes. The *V. parahaemolyticus* T3SS2α *vpa1362* genes formed a group supported by a significant bootstrap value (100), while the *V. cholerae* gene was separated and the T3SS2β gene formed the outgroup (Figure 2.10 B).

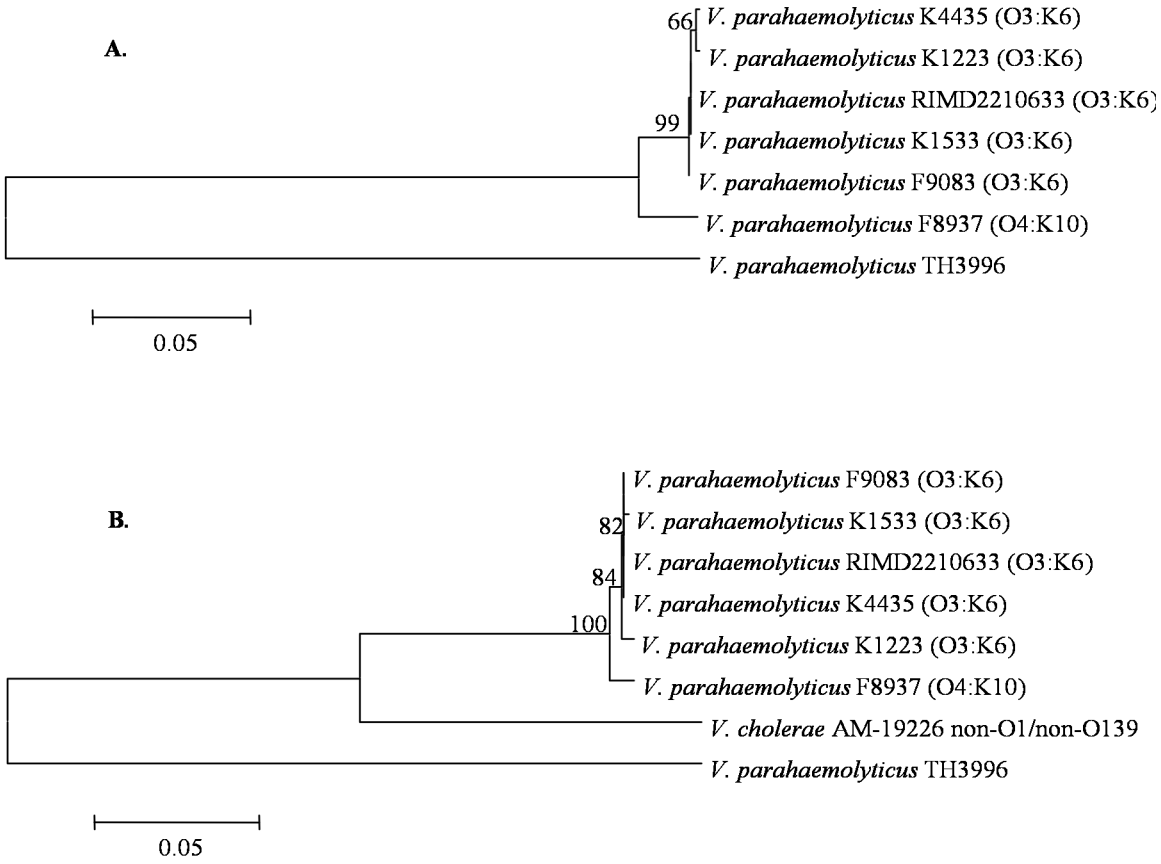


Figure 2.10. Neighbor-joining phylogenies of the nucleotide sequences of T3SS2 effector protein-encoding genes *vpa1346* (A) and *vpa1362* (B) of select *V. parahaemolyticus* clinical and environmental strains. The trees were constructed using the Kimura 2-parameter model and 1,000 bootstrap replications. Only bootstrap values ≥50 are shown. The scale bar indicates the evolutionary distance of 0.05 nucleotide changes.

Discussion

The increased frequency of *V. parahaemolyticus*-associated infections following the emergence of the clonal pandemic strains has generated a need for more research into the evolution of *V. parahaemolyticus* pathogenicity. A recent MLST study that analyzed seven housekeeping genes of globally distributed strains indicated *V. parahaemolyticus* has high genetic diversity (26); however, few studies have examined the localized population structure of *V. parahaemolyticus* strains in the environment. The examination of seven housekeeping genes provided greater resolution of the relationships of the clinical strains and resulted in the identification of two additional clonal complexes (26). In addition, previous studies have demonstrated that co-occurring *Vibrios* exhibit niche-associated genetic diversity (33).

In the current study, we analyzed the diversity of *V. parahaemolyticus* environmental strains for evidence of the emergence of disease-causing strains from environmental populations. We analyzed the diversity of four rather than seven housekeeping genes due to the high genetic diversity that has previously been demonstrated for *V. parahaemolyticus* environmental strains (26). In addition, previous studies have determined the evolutionary relationships of the non-clonal *V. parahaemolyticus* strains using three to four housekeeping genes (9, 14).

Previous studies have examined the diversity of geographically distributed strains (26), while in the current study we analyzed the diversity of a larger number of geographically co-occurring in addition to geographically separated strains. We examined the localized population structure of three geographically separated populations of *V. parahaemolyticus* strains from the Atlantic Ocean and the Gulf of Mexico. In

addition, we examined the temporal variation in population structure of *V. parahaemolyticus* strains isolated from GA in 2006 and 2007. The Atlantic Ocean sample sites in GA and NC may experience episodic mixing due to water transport by currents or the mobilization of strains by organisms and these sites may be considered inter-mixing populations. In contrast the Gulf of Mexico population should be significantly isolated from water exchange with the Atlantic sites that mixing may be limited to the transport of strains by animal hosts such as dolphins (10).

Overall, there were groups of related strains that possessed the virulence-associated genes suggesting that many of the pathogenic *V. parahaemolyticus* strains emerge from a common ancestor. Although the clonal pandemic strains appear to have evolved within a pathogenic lineage, there are several clinical strains that possess virulence-associated genes that appear to have evolved independently. Alternately, these strains may have been environmental strains that had horizontally-acquired virulence-associated genes. PFGE analysis of select *V. parahaemolyticus* clinical and environmental strains revealed unique patterns for all environmental strains examined. In contrast, the *V. parahaemolyticus* O3:K6 clonal pandemic strains and the clonal O4:K12 strains exhibited similar patterns. The dendrogram of the PFGE patterns exhibited a similar topology to the MLST trees. For example, the O3:K6 clonal pandemic strains and the O4:K12 strains formed groups within both the MLST and PFGE trees. In addition, the environmental strains AF34 and AF67 exhibited identical PFGE patterns; however, these strains exhibited relatedness to different strains in the MLST phylogeny. Additional PFGE analysis would be required to determine whether there is a conserved pattern for all of the strains within the MLST group that possessed AF34 and AF67.

Comparison of the phylogenetic analysis with the MGE content showed that MGEs are frequently lost and reacquired by strains. Strains of the biotype II *V. vulnificus* that causes infections in eels were shown to have horizontally acquired pathogenicity by a virulence plasmid (42). In addition, plasmid-encoded virulence has been previously characterized for *V. anguillarum* (19). The thermostable direct hemolysin, *tdh*, is associated with insertion sequences (37, 67) and has been identified on plasmid DNA (3) indicating a role of horizontal gene transfer in the movement of hemolysins.

Phylogenetic comparison of *V. parahaemolyticus* clinical and environmental strains revealed several strains that may be acquiring or losing pathogenicity genes. There were nine clinical strains that did not have the hemolysins or the T3SS2 genes; however, three of these strains formed a group that was related to the pandemic O3:K6 strains. These results indicate that these strains may have possessed and subsequently lost the genomic island VPaI-7 encoding *tdh* and the T3SS2 genes of the pandemic strains (34).

In addition, there were two clinical strains that lacked the virulence-associated genes that grouped together among the environmental strains. These strains were isolated from disease-associated samples; however, they may have been co-occurring with other *V. parahaemolyticus* strains that had the virulence-associated genes. Although a recent study indicated that multiple *V. parahaemolyticus* strains were shown to co-occur in disease-associated samples at least several of the strains that did not have hemolysins may have spontaneously lost them by the activity of insertion sequences (7, 37). Evidence of the loss of *tdh* was present for the *V. parahaemolyticus* clinical strain F8937 that grouped within the large group of mostly clinical strains that included the pre-1995

O3:K6 clinical strain ATCC 17802 and several environmental strains in addition to the sub-group of post-1995 O3:K6 pandemic strains. This suggests that F8937 may be emerging as a pathogen similar to the pandemic strains resulting from the acquisition of horizontally transferred genes such as the VPaI-7 genomic island and may have lost the *tdh* genes. The eight environmental strains within the large clinical group may likewise be emerging pathogens that could lead to disease outbreaks if they receive the necessary virulence-associated genes.

The identification of the *V. parahaemolyticus* environmental strain AF91 that exhibits remarkable genetic similarity to the *V. parahaemolyticus* clonal pandemic strains indicated the emerging pandemic clones from environmental populations. *V. parahaemolyticus* AF91 did not possess the virulence-associated genes; however, following the acquisition of the necessary genes by HGT this strain may emerge as a clonal pathogen. In contrast, this strain may represent a pandemic clone that has re-entered the environment and has begun to lose the virulence-associated genes.

In the current study, we have demonstrated the presence of *V. parahaemolyticus* environmental strains that exhibit significant relatedness to disease-causing strains indicating that emerging pathogens may exist in environmental populations. The identification of these intermediate strains indicates that HGT may play a significant role in the emergence of pathogenic *V. parahaemolyticus* strains that lead to outbreaks. Previous studies that have estimated the pathogenic potential of an environmental population by focusing on the presence of only the thermostable direct hemolysins may have significantly underestimated the pathogenic potential of environmental strains.

Further studies are necessary to determine the genetic repertoire required for invasion and establishment of a *V. parahaemolyticus*-associated infection. Identification of the full set of genes involved in the pathogenic mechanism of *V. parahaemolyticus* would allow further understanding of the evolution of disease-causing strains and their natural reservoir in the environment. In addition, the development of rapid methods for identification of *V. parahaemolyticus* strains and discrimination of the evolutionary relationships of disease-causing strains would facilitate the investigation of strains isolated from outbreak-related samples as outbreaks occur.

Acknowledgements

I would like to thank S. Chen, D. Silberger, and T. Lowe for laboratory assistance. Also, I would like to thank Jeanette Taylor for generating the TEM images.

Part 2. Identification of *Vibrio parahaemolyticus* By Whole-Cell Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry

Abstract

Vibrio parahaemolyticus is a pathogenic marine bacterium that is the main causative agent of seafood-borne gastroenteritis in the United States. An increase in the frequency of *V. parahaemolyticus*-related infections during the last decade has been attributed to the emergence of a clonal pandemic strain from India in 1995. The diversity of the O3:K6 clonal strains and serovariants has been examined using multiple molecular techniques including multilocus sequence typing (MLST), pulsed-field gel electrophoresis (PFGE), and group-specific PCR (GS-PCR) analysis. Matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) has become a powerful tool for rapidly distinguishing between related bacterial species. In the current study we demonstrate the development of a whole-cell MALDI-TOF MS method for the identification of *V. parahaemolyticus* from other *Vibrio* spp. We identified 29 peaks that were characteristic of the *V. parahaemolyticus* clinical and environmental strains examined in this study and may be developed as MALDI-TOF MS biomarkers for identification of *V. parahaemolyticus*. The majority of the *V. parahaemolyticus* strains identified had peaks with the greatest intensity within the *m/z* 6,000 to 8,000 region. The MALDI-TOF MS fingerprints of the *V. parahaemolyticus* strains examined were distinct from the other *Vibrios* examined including the closely related *V. harveyi* and *V. campbellii*. In addition, there was a peak identified at *m/z* 10,430 that was present in the MALDI-TOF MS spectra of all clonal O4:K12 strains analyzed. The results of this study demonstrate the first use of whole-cell MALDI-TOF MS analysis for the identification of *V. parahaemolyticus* from related *Vibrios.*

Introduction

Vibrio parahaemolyticus is one of several pathogenic *Vibrios* that occurs in coastal environments and is the leading cause of seafood-borne gastroenteritis in the United States (1). Following the emergence of the *V. parahaemolyticus* O3:K6 clonal pandemic strain in India in 1995 there has been a rise in the number of reported *V. parahaemolyticus*-associated illnesses each year making it a pathogen of increasing concern (1, 17). The *V. parahaemolyticus* pandemic strains may have emerged following the acquisition of several genomic islands that are characteristic of the O3:K6 clones and related serovariants (28, 34).

The initial identification of *V. parahaemolyticus* is often conducted by growing strains on the selective media thiosulfate citrate bile salts sucrose (TCBS) agar (22, 39). TCBS provides an initial selective measure to separate *Vibrio* spp. from the background of diverse marine bacteria; however, it does not allow distinction of *V. parahaemolyticus* from closely related *Vibrios* such as *V. harveyi* and *V. campbellii.* Additional molecular analyses are required to positively identify *V. parahaemolyticus* strains. Recent studies have increasingly employed molecular techniques to examine the diversity of *V. parahaemolyticus* strains. Molecular identification of *V. parahaemolyticus* routinely involves serotyping (18, 22), pulsed-field gel electrophoresis (PFGE) (36, 58), group-specific PCR (GS-PCR) (7), multiplex PCR (66), multilocus sequence typing (MLST) (14, 26), and comparative gene arrays (71). Although many of these techniques can be used to distinguish *V. parahaemolyticus* from related *Vibrio* spp. they do not provide enough resolution to examine the evolution of the clonal pandemic strains. The development of a technique to distinguish *V. parahaemolyticus* from other *Vibrio* pathogens would greatly aid the rapid identification of strains involved in disease

outbreaks when time is critical. The emergence and spread of the clonal *V. parahaemolyticus* pandemic strains belonging to the O3:K6 serotype and related serovars has heightened the need for a rapid and discriminatory technique to monitor the evolutionary relationships of these strains (7).

Recent studies have demonstrated that whole-cell MALDI-TOF MS analysis is a powerful tool for the rapid identification of bacterial species. Recent studies have used MALDI-TOF MS to characterize *Streptococcus* sp. (74), *Salmonella* spp. (20), *Mycobacteria* sp. (59), and *Listeria* spp. (5). These studies demonstrated the use of whole-cell MALDI-TOF MS analysis to generate a highly reproducible profile that is characteristic of bacterial species and can be used to distinguish among strains. Whole-cell MALDI-TOF MS analysis allows the accurate and reproducible generation of bacterial fingerprints that may be analyzed for the presence of biomarker peaks present in a species or clonal group (5, 41, 59, 74). The whole-cell MALDI-TOF MS procedure involves growing cells under optimized conditions and resuspending the cells in a matrix directly on a MALDI-TOF MS plate. The MALDI-TOF MS fingerprint generated may be analyzed similarly to other current molecular fingerprinting techniques such as pulsed-field gel electrophoresis (PFGE). PulseNet International is a network that uses PFGE fingerprints to identify and monitor food-borne pathogens such as *V. parahaemolyticus* (36, 58). PFGE has been a reliable method for identifying the clonally-related *V. parahaemolyticus* pandemic O3:K6 strains and other related serovariants that typically have identical PFGE fingerprints (44); however, methods that provide greater resolution among clonal strains would be required for investigating the evolution of these strains. MALDI-TOF MS has provided greater resolution for analyzing the diversity of

Salmonella spp. at the subspecies level compared to other molecular techniques (20). In this regard, MALDI-TOF MS may be used to examine the evolution of the *V. parahaemolyticus* O3:K6 clonal pandemic strains and serovariants that are associated with many of the disease outbreaks (1).

In this study, we developed a method for whole-cell MALDI-TOF MS analysis of *V. parahaemolyticus* strains. MALDI-TOF MS analysis was used to distinguish *V. parahaemolyticus* from six other closely related *Vibrio* spp. and identify potential *V. parahaemolyticus*-specific biomarker peaks. We demonstrate that MALDI-TOF MS analysis is reliable for distinguishing *V. parahaemolyticus* from closely-related *Vibrio* spp. including *V. harveyi* and *V. campbellii.* MALDI-TOF MS analysis demonstrated that clonal strains exhibit detectable variation in their MALDI-TOF MS fingerprints. In addition, MALDI-TOF MS analysis of a quorum sensing mutant and a mismatch repair mutant produced a unique spectrum compared to the wild-type strain indicating that MALDI-TOF MS may be used to monitor protein evolution resulting from relatively few genetic changes. This is the first study that demonstrates whole-cell MALDI-TOF MS analysis may be used to characterize *Vibrio* spp. and to identify potential *V. parahaemolyticus*-specific peaks that could be developed as biomarkers. Further application of the results obtained in this study would involve the development of a database of MALDI-TOF MS spectra from diverse *Vibrio* spp. for the application of whole-cell MALDI-TOF MS analysis for rapid identification of *Vibrios*.

Materials and methods

Bacterial strains, media, and growth conditions. The *V. parahaemolyticus* clinical strains examined in this study were isolated from illness-related human and food samples of diverse outbreaks by the Centers for Disease Control and Prevention in Atlanta, GA. The *V. parahaemolyticus* environmental strains analyzed were isolated from sediment, water, and oyster samples from GA, FL, and NC as previously described (4, 16, 30). All *V. parahaemolyticus* strains were grown in Heart Infusion broth with 2% NaCl supplemented with 2% agar for plates.

Table 2.7. Bacterial strains examined in this study

Bacteria	Strain	Year	Serotype	Location	Sample type	Source
V. parahaemolyticus	RIMD2210633	1996	O3:K6	Japan	Clinical	(43)
	K1223	2004	O3:K6	CO	Clinical	(46)
	F9083	2002	O3:K6	AZ	Clinical	"
	K4435	2006	O3:K6	GA	Clinical	"
	17802	1965	O3:K6	Japan	Clinical	(24)
	F8023	2001	O4:K12	GA	Clinical	"
	K1461	2004	O4:K12	MA	Clinical	"
	F5052	1997	O4:K12	WA	Clinical	"
	K3566	2006	O4:K63	LA	Clinical	"
	K4250	2006	O4:K63	NY	Clinical	"
	K4981	2007	O1:Kuk	OK	Clinical	"
	K5067	2007	O1:K56	SD	Clinical	"
	K5330	2007	O5:Kuk	TX	Clinical	"
	22702	1998	O5:Kuk	GA	sediment	(16)
	SG176	2006	O5:Kuk	GA	water	(30)
	SG258	2006	O1:Kuk	GA	sediment	"
	AF2	2006	O4:K9	FL	oyster	"
	AF67	2006	O5:K37	FL	sediment	"
	J-C2-15	1998	O1:Kuk	NC	sediment	(4)
	S-M2-3-B3	1998	O5:K17	NC	sediment	"
V. cholerae	O395		N/A	N/A	N/A	
V. harveyi	ATCC BAA-1116		"	"	"	
V. fischeri	ES114		"	"	"	(62)
V. campbellii	09022		"	"	"	(31)
V. fluvialis	0908		"	"	"	"
V. mediterranei	23023		"	"	"	"
E. coli K12	MG1655		"	"	"	(8)

Serotyping. The serotypes of the *V. parahaemolyticus* clinical and environmental strains were determined as previously described (18, 22). A slide agglutination test was performed using anti-O and anti-K antisera (Denka; Seiken Corp., Tokyo, Japan).

MALDI-TOF MS. The MALDI-TOF MS spectra were generated on an Applied Biosystems MALDI-TOF mass spectrometer Voyager DE STR operating in the linear, delay extraction, and positive ion modes. The laser intensity (N_2, 337 ns) was set above the ion generation threshold. The low mass gate was *m/z* 2,400 with a delay time of 750 ns. The accelerating voltage was 25,000 V and the grid voltage was 95% of the accelerating voltage.

Vibrio strains were grown for approximately 24 hours in 3 ml of HI broth with shaking (200 rpm) at 30˚C. Following incubation, 1 ml of each culture was added to a microcentrifuge tube and centrifuged for 2 min at 12,000 rpm. The supernatant was decanted and cells were washed by resuspending in 1 ml of 0.85% NaCl. Then the cell suspension was centrifuged for 2 min at 12,000 rpm and the supernatant was decanted. Cells were washed two times with 1 ml of 50% ethanol and centrifuged for 2 min at 12,000 rpm. Following two washes with ethanol the cells were resuspended to 0.2 mg/µl in 1% trifluoroacetic acid (TFA).

The optimal MALDI-TOF MS matrix for preparation of the *Vibrio* strains analyzed consisted of 10 mg/ml sinapinic acid, 50% acetonitrile, 50% water, and 0.1% trifluoroacetic acid. Cytochrome c was added to the matrix at a concentration of 2 pmol/µl as an internal standard. Strains were spotted on each MALDI-TOF MS plate by mixing 0.5 µl of bacterial suspension and 0.5 µl of MALDI-TOF MS matrix directly on the plate. The plate was allowed to air dry before loading onto the machine. Each strain

was spotted on the plate in triplicate and MALDI-TOF MS spectra were generated for each strain by performing three independent runs. MALDI-TOF MS spectra were generated on a MALDI-TOF mass spectrometer Voyager DE STR (Applied Biosystems Inc.; Foster City, CA). The mass data were generated by 300 laser shots in a *m/z* range of 2500-15000 Daltons. The resulting mass data were calibrated using the doubly (*m/z* 6181.0) and singly (*m/z* 12361.0) charged peaks of the cytochrome c internal standard. MALDI-TOF MS data analysis. A single representative spectrum of the three replicates for each strain from a single run was analyzed. The peak intensities were adjusted to 0.5% or greater relative to the most intense peak using the software provided with the instrument. A total of three separate runs for each strain analyzed were combined to generate a composite list of peaks. The peaks that occurred within the same confidence interval denoted by the upper and lower mass bounds were averaged to generate a single data point. Peaks not present in two or more independent runs for each strain were excluded from the analysis. The combined spectra of each of the strains were then compared and sorted. A binary matrix of 0 and 1 was assigned for each unique peak identified among all strains examined to indicate the presence or absence of a peak for a particular strain. Peaks that were identified in only one of the strains examined were excluded.

The composite list of peaks for each strain analyzed were compared and used to generate an absence/presence matrix for cluster analysis denoted by 0 and 1, respectively. The similarity matrix was analyzed using PAST (v. 1.34) (http://folk.uio.no/ohammer/past/doc1.html) as previously described (74). The Jaccard

similarity coefficient with single linkages was used for cluster analysis of MALDI-TOF MS data.

Biomarker peak identification. Representative peaks that were present in the MALDI-TOF MS profiles of multiple species and strains for possible use as biomarkers for species and strain level identification were identified using the ExPASy TagIdent tool (25). Peaks differing between the Δ*opaR* and RIMD2210633 MALDI-TOF MS spectra and the Δ*mutS* and ATCC 17802 spectra were examined by comparison with predicted proteins of the RIMD2210633 genome using TagIdent (25).

PFGE. Pulsed-field gel electrophoresis of the *V. parahaemolyticus* clinical and environmental strains was performed according to the PulseNet protocol (58). Restriction endonuclease patterns were generated using the restriction endonucleases SfiI and NotI. PFGE patterns were analyzed with Bionumerics (v. 5.0) by comparing to the XbaI restricted control strain *Salmonella branderup* H1981.

MLST. Multilocus sequence typing of seven housekeeping genes of *V. parahaemolyticus* clinical and environmental strains was performed as previously described (26). PCR amplification was performed with NEB Phusion High Fidelity polymerase with the GC buffer. All PCR amplicons were gel-purified on a 0.7% SeaKem LE agarose gel (Lonza; Switzerland) containing ethidium bromide. The expected products were excised from the gel and extracted using a Sigma GenElute gel extraction kit (Sigma Aldrich; St. Louis MO). Sequencing was performed using M13 universal primers at the Georgia Institute of Technology genomics facility.

Sequence alignments and a concatenation of the MLST data were performed in MEGA (v. 4.1) (40). A neighbor-joining tree was constructed from the concatenated

nucleotide sequences with the Kimura 2-parameter correction model and 1,000 bootstrap replications. The concatenated tree was constructed from *recA* (642 nt), *gyrB* (618 nt), *pntA* (414 nt), *pyrC* (531 nt), *dtdS* (498 nt), *dnaE* (587 nt), and *tnaA* (470 nt) nucleotide sequences.

Accession numbers. The seven housekeeping gene sequences (*recA, gyrB, pyrC, pntA, dtdS, dnaE,* and *tnaA*) for each *V. parahaemolyticus* strain examined in this study were deposited in GenBank under the accession numbers FJ577370-FJ577500. The *recA* of K1223 and 22702 were previously submitted under the accession numbers EU652262 and EU018456, respectively.

Results

Whole-cell MALDI-TOF MS identification of *Vibrios*. The whole-cell MALDI-TOF MS method was used to generate spectra that were highly reproducible as shown for *V. parahaemolyticus* RIMD2210633 (Figure 2.11). The spectra shown in Figure 2.11 represent six individual runs of *V. parahaemolyticus* RIMD2210633. The same prominent peaks with similar intensity were detected for different spectra generated for a single strain such as that shown for RIMD2210633 (Figure 2.11). The peak intensity varied among the strains examined; however, there was relatively little within strain variation between separate runs that were performed on different days with different overnight cell cultures (Figure 2.11).

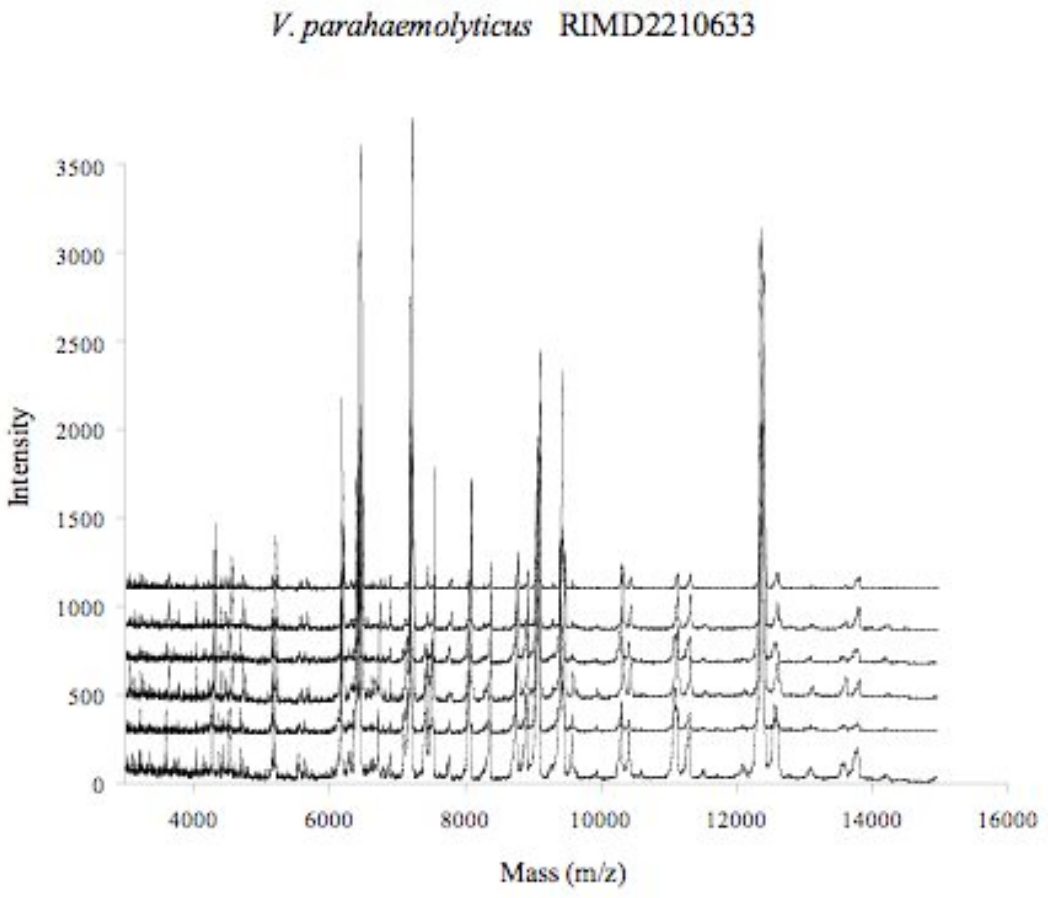
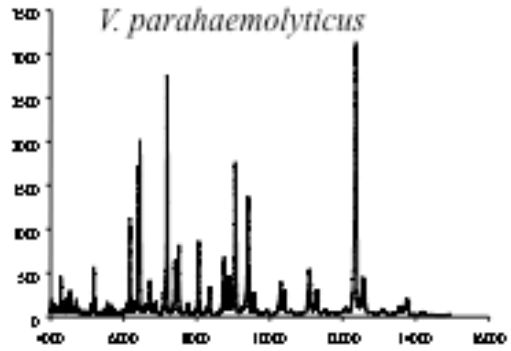
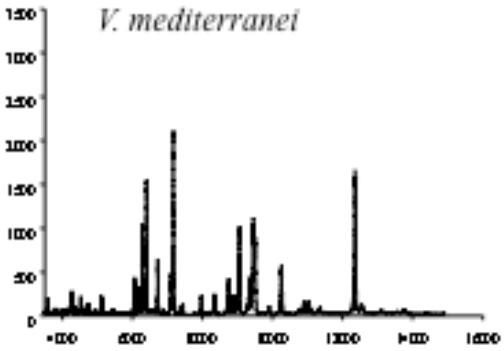
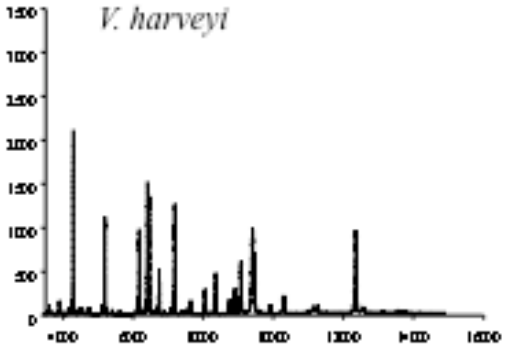


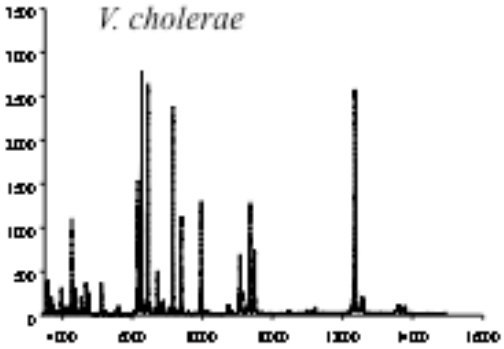
Figure 2.11. MALDI-TOF MS spectra of *V. parahaemolyticus* RIMD2210633 showing the reproducibility of spectra. Each MALDI-TOF MS profile represent an independently generated data set.

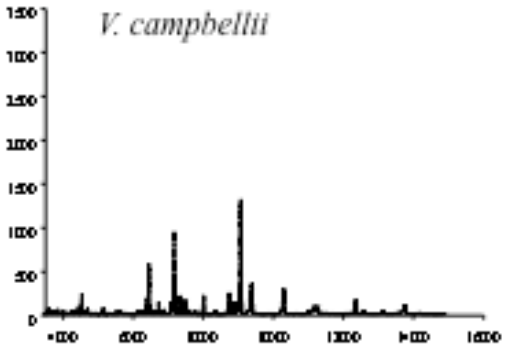
Analysis of diverse *Vibrios* by whole-cell MALDI-TOF MS generated diverse spectra that could be used to distinguish *V. parahaemolyticus* from other species including the closely related *Vibrios* such as *V. harveyi* and *V. campbellii.* The non-*V. parahaemolyticus* species examined also included an environmental strain *Vibrio* sp. 0908, which is most-related to *V. fluvialis* and *V. furnissii* (31) and the environmental strain *Vibrio* sp. 23023, which is most related to *V. mediterranei* (31) (Table 2.7). In addition, spectra were generated for the distantly related species *V. cholerae* and *V. fischeri* (Table 2.7). The spectra generated for each of the seven *Vibrio* spp. analyzed were unique (data not shown). Representative spectra are shown for *V. parahaemolyticus, V. harveyi, V. campbellii, V. fischeri, V. fluvialis, V. mediterranei,* and *V. cholerae* (Figure 2.12). The MALDI-TOF MS spectra of the non-*V. parahaemolyticus* species examined were significantly different from the *V. parahaemolyticus* spectra analyzed (Figure 2.12).

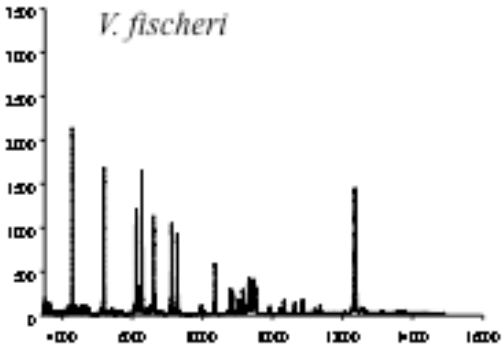

V. parahaemolyticus

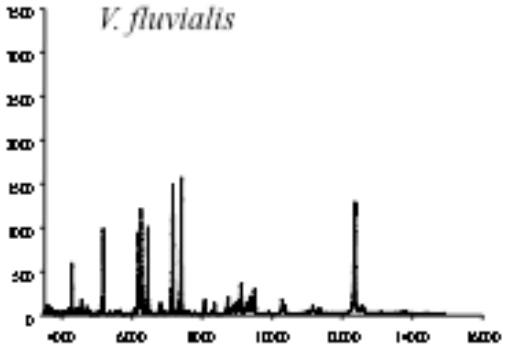

V. mediterranei


V. harveyi


V. cholerae


V. campbellii


V. fischeri


V. fluvialis

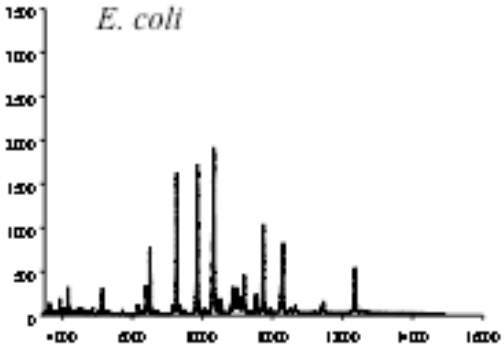

E. coli

Figure 2.12. Whole-cell MALDI-TOF MS spectra of eight *Vibrio* spp. including the *V. parahaemolyticus* O3:K6 clonal pandemic strain RIMD2210633. The x-axis shows the mass intensity of the peaks on a *m/z* scale of 4,000 to 16,000. Each spectrum is a representative run from a total of three independent runs perfomed for each strain.

The MALDI-TOF MS spectra of each species examined had a unique fingerprint that could be visually compared to assess differences among *Vibrios*. Each spectrum had a single prominent peak at *m/z* 12,361 from the cytochrome c internal standard. The majority of the peaks in the *V. harveyi* spectrum were present in two clusters at approximately the *m/z* 6,000 to 7,000 range and from *m/z* 8,200 to 9,500 with additional peaks present at approximately *m/z* 4,000 and 5,000 (Figure 2.12). The *V. campbelli* likewise formed two cluster at approximately 6,000 to 7,000 and 8,200 to 9,500 but lacked the additional peaks present at 4,000 and 5,000 (Figure 2.12). In addition, the *V. campbellii* spectrum had several small peaks around *m/z* 6,000 that were not present in the *V. harvei* spectrum (Figure 2.12). The *V. fischeri* spectrum was the most unique with the highest intensity peaks present at approximately *m/z* 4,000 and 5,000 and a decreasing peak intensity correlated with increasing peak mass (Figure 2.12). There also were peaks present in the *V. fischeri* spectrum that formed clusters at *m/z* 6,000 to 7,000 and 8,100 to 9,500; however, the number and intensity of the peaks varied significantly from the other strains examined (Figure 2.12). The *V. cholerae* spectrum had numerous low intensity peaks from *m/z* 3,500 to 4,500 and the two clusters of peaks located at *m/z* 6,000 to 7,500 and 9,000. In addition, *V. cholerae* had a prominent peak at *m/z* 8,000 that was also present in the *V. parahaemolyticus* spectra (Figure 2.12).

MALDI-TOF MS analysis of *V. parahaemolyticus* clinical and environmental strains. To determine whether whole-cell MALDI-TOF MS can be used to distinguish between *V. parahaemolyticus* strains including the clonal pandemic strains, we analyzed *V. parahaemolyticus* clonal O3:K6, clonal O4:K12, non-clonal clinical, and non-clonal environmental strains.

There were visual differences in the MALDI-TOF MS spectra of post-1995 clonal O3:K6 strains examined compared to a pre-1995 O3:K6 strain (Figure 2.13). An extra peak was present within the *m/z* 4,000 to 8,000 range of the spectra from the pre-1995 O3:K6 strain ATCC 17802 that was not present in the spectra of the post-1995 clonal O3:K6 strains examined (Figure 2.13). In addition, there were several prominent peaks present at *m/z* 7,000 in the spectra of the post-1995 O3:K6 strains that were absent from the spectrum of the pre-1995 O3:K6 strain (Figure 2.13).

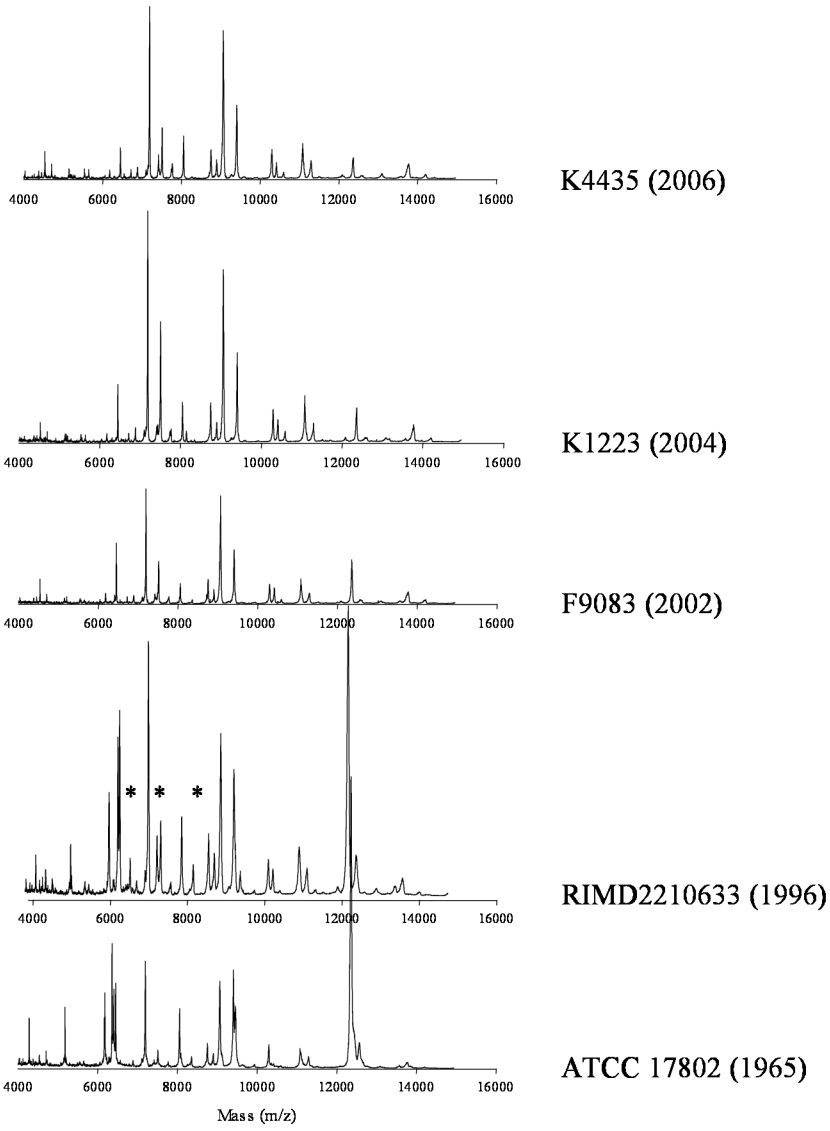


Figure 2.13. MALDI-TOF MS fingerprints of *V. parahaemolyticus* post-1995 clonal pandemic O3:K6 strains (K4435, K1223, F9083, RIMD2210633) and a pre-1995 non-clonal O3:K6 strain ATCC 17802. The x-axis indicates the mass of each peak in the range of 4,000 to 16,000 Da. The y-axis of each individual plot indicates the peak intensity and has a scale of 0 to 3,500 (not shown). The year that each strain was isolated is indicated in parentheses. Peaks present in the RIMD spectra that were absent in the 17802 spectra are indicated.

In addition, we examined the differences in the MALDI-TOF MS fingerprints of clonal O4:K12 strains (Figure 2.14). The spectra of the more recently isolated strains F8023 and K1461 were nearly identical, except for a few additional peaks present in the 6,000 to 8,000 range of the K1461 spectrum (Figure 2.14). In contrast, there were additional peaks detected in the 4,000 to 6,000 range of F5052 that were not present in the spectra of F8023 and K1461 (Figure 2.14). The majority of the peaks were present in the 6,000 to 13,000 range except for three additional peaks of F5052 that were present in the 4,000 to 5,000 range (Figure 2.14). There was a peak detected at m/z 10,428 to 10,432 that was present in the three O4:K12 strains and no other strains examined indicating this peak may be a biomarker for the clonal O4:K12 strains.

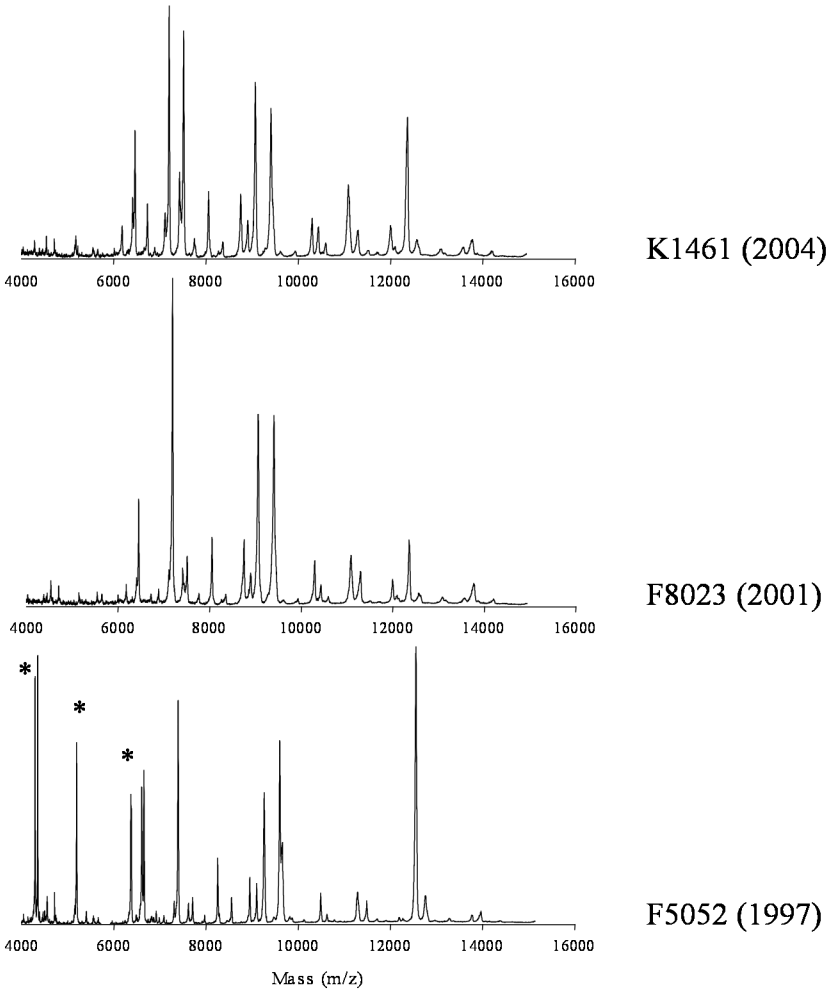


Figure 2.14. MALDI-TOF MS fingerprints of the clonal O4:K12 strains K1461, F8023, and F5052. The x-axis indicates the mass of each peak in the range of *m/z* 4,000 to 16,000. The y-axis of each individual plot indicates the peak intensity and has a scale of 0 to 3,500 (not shown). The year that each strain was isolated is indicated in parentheses. Peaks present in one spectrum that were absent in another are indicated.

The *V. parahaemolyticus* environmental strains examined exhibited unique MALDI-TOF MS fingerprints (Figure 2.15). There were visual similarities present among strains isolated from the same location such as the J-C2-15 and S-M2-3-B3 strains isolated from NC (Figure 2.15).

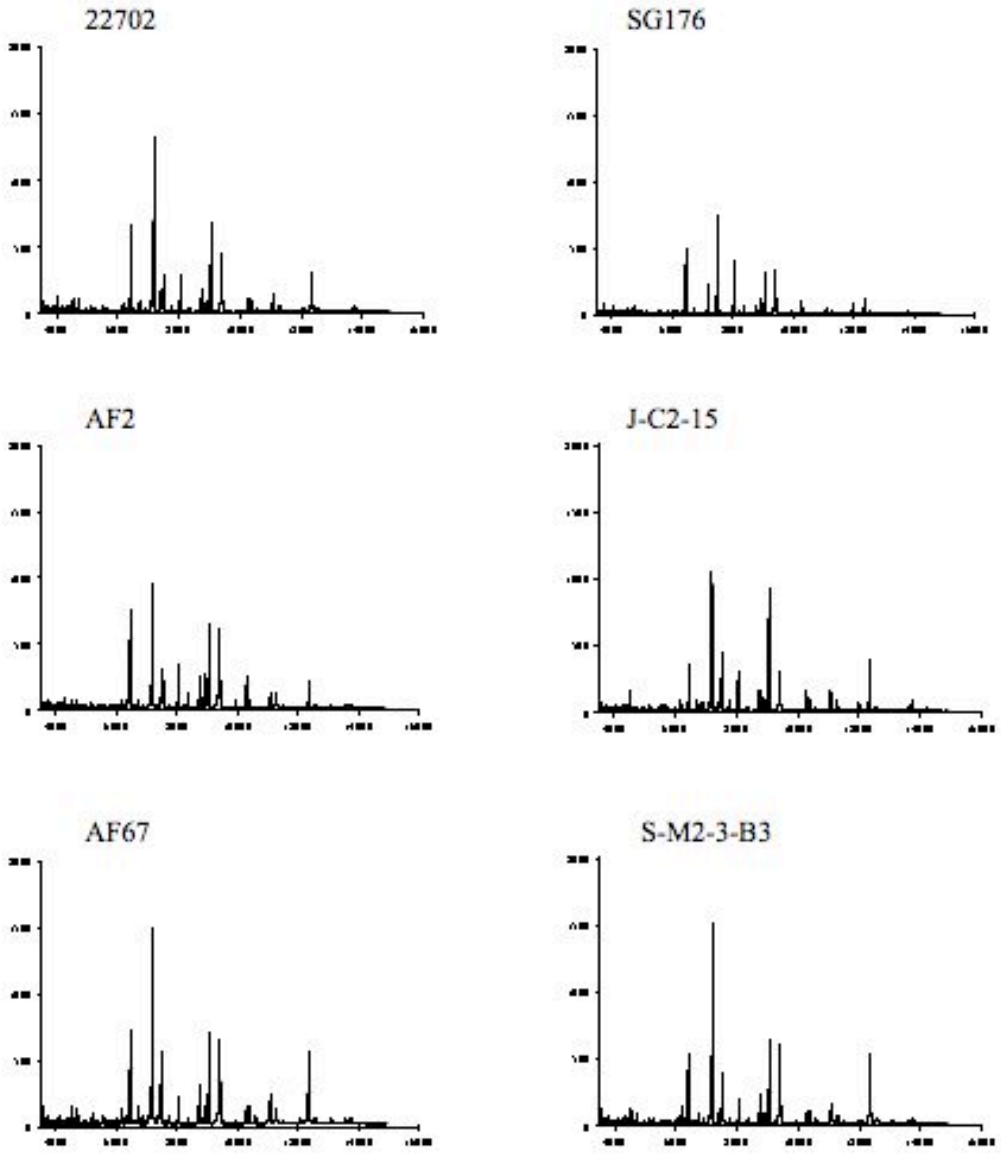


Figure 2.15. MALDI-TOF MS fingerprints of environmental strains isolated from sediment and water samples from GA, FL, and NC. The x-axis indicates the mass of each peak in the range of 4,000 to 16,000 Da. The y-axis of each individual plot indicates the peak intensity and has a scale of 0 to 3,500 (not shown). The year that each strain was isolated is indicated in parentheses.

Cluster analysis performed in PAST using the Jaccard similarity coefficient was used to generate a dendrogram that was compared to phylogenies constructed from MLST and PFGE data (Figure 2.16). The MALDI-TOF MS cluster analysis was performed using all peaks generated and the dendrogram shown is constrained to better illustrate differences in the level of MALDI-TOF MS fingerprint divergence among the *V. parahaemolyticus* strains (Figure 2.16). The *V. parahaemolyticus* strains examined formed a distinct group from the other *Vibrios* examined and *E. coli* (Figure 2.16). The *V. parahaemolyticus* strains exhibited from 42 to 61% similarity and approximately 18% similarity to *V. harveyi, V. campbellii*, and *V. cholerae* (Figure 2.16). In contrast, *V. parahaemolyticus* had only 12 to 14% similarity to *V. mediterranei, V. fluvialis,* and *V. fischeri* (Figure 2.16). Of the seven environmental strains examined there were five that exhibited greater divergence (Figure 2.15). The unconstrained tree topology exhibited a strongly correlated *V. parahaemolyticus* group that had 40 to 70% similarity (data not shown). *V. harveyi* and *V. campbellii* exhibited 20% and 30% similarity while the remaining species exhibited <20% similarity indicating *V. harveyi* and *V. campbellii* MALDI-TOF MS data was more similar to the *V. parahaemolyticus* data (data not shown).

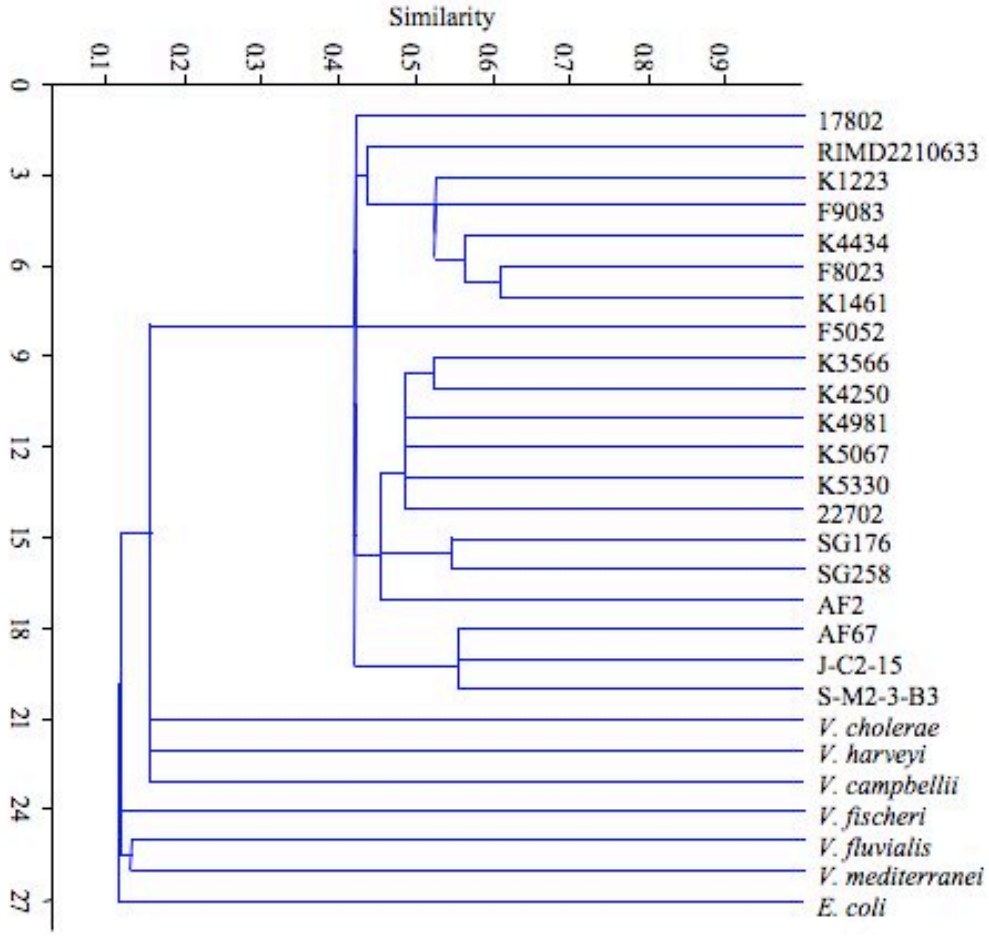


Figure 2.16. Cluster analysis of *Vibrio* spp. examined using the Jaccard similarity coefficient with single linkage associations. The topology of the tree is constrained and the relationships of the strains are represented by comparing levels of similarity.

Identification of *V. parahaemolyticus* by PFGE and MLST. Analysis of the PFGE patterns of the 20 *V. parahaemolyticus* clinical and environmental strains examined revealed unique patterns for nearly all of the strains examined (Fig. 6). The O3:K6 clonal pandemic strains and O4:K12 strains exhibited similar PFGE profiles as indicated by the formation of groups of these strains in the dendrogram. The PFGE profiles of the remaining clinical and environmental strains did not exhibit relatedness based on serotype or isolation source. An exception were the PFGE profiles of the environmental strains AF2 and J-C2-15 that were isolated from FL and NC, respectively.

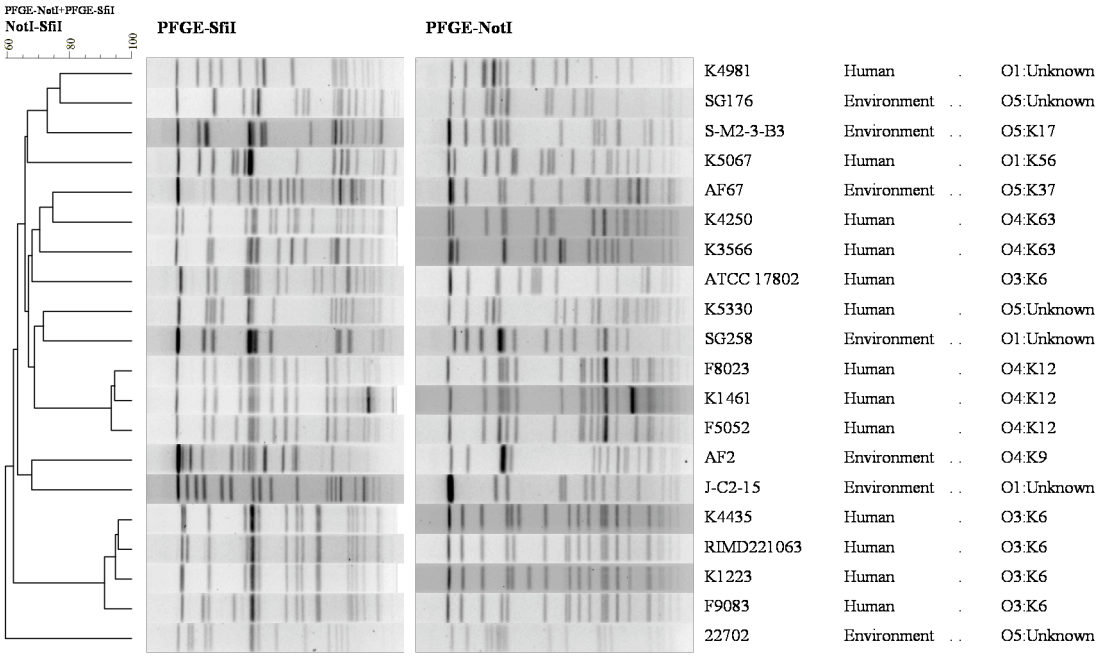


Figure 2.17. Dendrogram and PFGE patterns of SfiI- and NotI-digested DNA from the *V. parahaemolyticus* clinical and environmental strains examined.

Another molecular method frequently used to identify *V. parahaemolyticus* that we compare in this study to MALDI-TOF MS indentification is MLST. To determine whether whole-cell MALDI-TOF MS analysis is reliable for distinguishing *V. parahaemolyticus* from other *Vibrios* and investigating the evolutionary relationships of *V. parahaemolyticus* strains, we performed an MLST analysis of seven housekeeping genes on all strains examined. Individual nucleotide trees of each of the seven genes analyzed revealed different topologies indicating these genes are under different levels of selection (data not shown). The formation of two clonal groups and the remainder of the *V. parahaemolyticus* strains comprising a larger group was consistent for each tree (data not shown) and for the concatenated tree. The clonal groups consist of *V. parahaemolyticus* strains of the O3:K6 and O4:K12 serotypes (Figure 2.18). An exception for the overall concatenation was the presence of strain 22702 in a group with the O3:K6 strains (Figure 2.18). *V. parahaemolyticus* 22702 was isolated from a sediment sample from a GA saltmarsh and has the O5:Kuk serotype. Analysis of the individual gene trees of each of the seven housekeeping genes showed 22702 was related to other environmental strains for six of the seven genes. In the *recA* tree *V. parahaemolyticus* 22702 formed a group with the O3:K6 strains (data not shown). The relatedness of 22702 to the O3:K6 strains was also observed in the seven-gene concatenated tree (Figure 2.18).

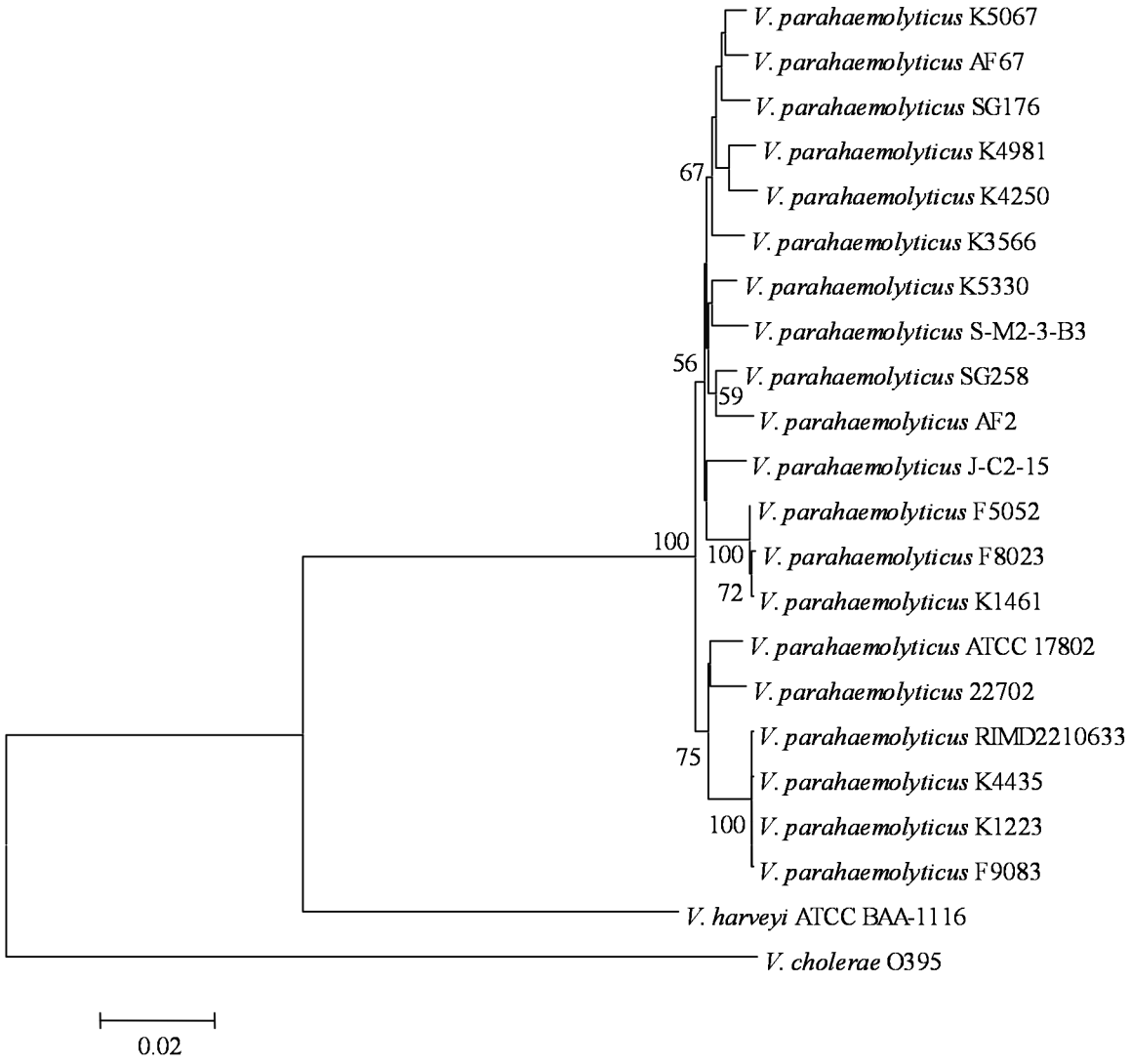


Figure 2.18. A neighbor-joining tree constructed with the concatenated nucleotide sequences (3,760 nt) of seven housekeeping genes (*recA, gyrB, pntA, pyrC, dtdS, dnaE,* and *tnaA*). The tree was constructed using the Kimura 2-parameter model and 1,000 bootstrap replications. Only bootstrap values ≥50 are shown. The scale bar represents the evolutionary distance of 0.02 nucleotide changes.

Identification of biomarker peaks and differentiation of wild-type and mutant strains by MALDI-TOF MS analysis. There were a total of 72 conserved peaks that were shared by at least two or more strains or species analyzed. Of the conserved peaks there were 29 that were detected only in the spectra of *V. parahaemolyticus* strains (Table 2.8). In addition, there was a peak present at m/z 6,409 that was detected in the spectra of 15 of the 20 *V. parahaemolyticus* strains examined as well as *V. campbellii, V. mediterranei,* and *E. coli* indicating this peak may be a biomarker of gamma-proteobacteria.

Identification of the proteins corresponding to the potential biomarker peaks was performed by searching the sequenced *V. parahaemolyticus* RIMD2210633 genome using the TagIdent tool of ExPASy (25). The majority of the peaks identified by comparing with *V. parahaemolyticus* RIMD2210633 predicted proteome were ribosomal proteins (Table 2.8). In addition, there were peaks that corresponded to regulatory proteins such as Hfq, and RpoZ and numerous hypothetical proteins (Table 2.8). Frequently, there were multiple predicted proteins corresponding to certain peaks and certain proteins that corresponded to multiple peaks within a certain size range (Da).

Table 2.8. Identification of MALDI-TOF MS peaks

Peak Mass (Da)	Predicted Genes	Predicted Proteins	Species present
3229.70	0		Vp
3586.65	0		Vc, Vfl, Vm
3601.14	0		Vp
3759.05	0		Vp
4029.91	0		Vp
4277.4-4293.8	VP0278	50S ribosomal protein L36	Vc, Vfl, Vf, Vh, Vp
4380.29	0		Vp
4452.72	0		Vp
4534.53	0		Vp, Vm
4541.05	0		Vp, Vca
4706.28	0		Vp
4725.96	0		Vp, Vf, Vm
5126.34	0		Vc, Vm, Vf
5150.3-5209.2	VP0005	50S ribosomal protein L34	Vp, Ec, Vh, Vf, Vfl
5541.76	0		Vp
5648.74	0		Vp
6005.42	0		Vp
6169.24	VP2058	50S ribosomal protein L32	Vp, Vca, Vc
6276.5-6307.7	VP0082	transmembrane protein	Vm, Vf, Vc, Vp
6322.21	0		Vp, Vf
6367.59	0		Vp
6395.2-6409	0		Vh, Vp, Vm, Vca, Ec
6430.33	0		Vp
6453.75	0		Vc, Vp, Vfl, Vca, Vh
6613.51	VP0275	50S ribosomal protein L30	Vp, Vf
6723.64	0		Vp
6738.19	0		Vca, Vh, Vc, Vm
6887.98	0		Vp, Vm, Vca, Vc
7103.7-7112.8	VP2546	CsrA	Vp, Vm, Vca, Vh
7123.9-7176.2	VP2546	CsrA	
	VP2529	zinc-binding protein	
	VP0265	50S ribosomal protein L29	

7195.5-7238.5	VP2529	zinc-binding protein	Vp, Vh
	VP0265	50S ribosomal protein L29	
7402.2-7405.9	VP2335	hypothetical protein	Vc, Vp, Vfl
	VP1281	50S ribosomal protein L35	
7415.7-7419	VP2335	hypothetical protein	Vp
7441.78	0		Vp, Vm
7513.33	0		Vp
7540.14	0		Vp
7605.64	0		Vca, Vp
7745.46	0		Vp
7771.33	0		Vp
8055.8-8106.9	VP2791	hypothetical protein	Vp, Vh, Vfl, Vm
	VP0255	50S ribosomal protein L31	
8356.9-8371.6	VP2053	Acyl carrier protein	Vm, Vfl, Vp, Vca, Vh, Vf
8711.92	VP3074	ATP synthase subunit c	Vp
	VP2780	SlyX	
	VP0098	Sec-independent protein translocase protein tatA/E homolog	
8737.71	VP2780	SlyX	Vc, Vfl
	VP0098	Sec-independent protein translocase protein tatA/E homolog	
8754.42	VP0098	Sec-independent protein translocase protein tatA/E homolog	Vp, Vm, Vca
8894.9-8902.2	VP2738	30S ribosomal protein S18	Vp, Vm, Vca, Vh, Ec
	VP2545	Probable oxaloacetate decarboxylase gamma chain	
	VP0688	Exodeoxyribonuclease 7 small subunit	
8964.15	VP0185	50S ribosomal protein L28	Vp
	VP2545	Probable oxaloacetate decarboxylase gamma chain	
	VP0688	Exodeoxyribonuclease 7 small subunit	
9023.5-9041.4	VP2545	Probable oxaloacetate decarboxylase gamma chain	Vp, Vca, Vfl
	VP2533	30S ribosomal protein S16	
	VP0185	50S ribosomal protein L28	
9066.4-9088.5	VP2533	30S ribosomal protein S16	Vm, Vh, Vp, Ec, Vca, Vc

	VP0185	50S ribosomal protein L28	
9121.9-9280.4	VP0329	50S ribosomal protein L27	Vc, Vfl, Ec, Vp
	VP2533	30S ribosomal protein S16	
	VP0185	50S ribosomal protein L28	
9367.9-9408.5	VP0246	Cell division protein zapB	Vc, Vca, Vfl, Vp
	VP0025	Sulfurtransferase tusA	
9445.5-9463.5	VP0531	30S ribosomal protein S20	Vp, Vf, Vh
	VP0246	Cell division protein zapB	
	VP0025	Sulfurtransferase tusA	
9508.51	VP0266	30S ribosomal protein S17	Vfl, Vm
	VP0531	30S ribosomal protein S20	
9909.88	VP2817	Hfq	Vm, Vh
	VP2331	50S ribosomal protein L31 type B	
9932.94	VP2817	Hfq	Vp
	VP2331	50S ribosomal protein L31 type B	
	VP2582	hypothetical protein	
	VP0160	RNAP omega subunit rpoZ	
10258.97	VPA0057	hypothetical protein	Vf, Vm
	VP2852	GroES protein 1	
	VP1627	Acylphosphatase	
	VP0718	hypothetical protein	
10297.26	VPA0057	hypothetical protein	Vp, Vf, Vca, Vh, Ec
	VP0718	hypothetical protein	
	VP2852	GroES protein 1	
10415.6-10430.5	VP1210	50S ribosomal protein L25	Vp
	VP0261	30S ribosomal protein S19	
10594.22	VPA0321	hypothetical protein sigma 70 region	Vp
	VPA0286	GroES protein 2	
	VP2627	Probable Fe(2+)-trafficking protein	
	VP2619	YggU-like hypothetical protein	
	VP2029	IHF-beta	
11040.5-11096	VP2885	DNA-binding protein fis	Vp, Vm
	VP2739	Primosomal replication protein n	
	VP0259	50S ribosomal protein L23	
	ureA	Urease subunit gamma	
11105.6-11144.9	VP2885	DNA-binding protein fis	Vp, Vh, Vfl
	VP2739	Primosomal replication protein n	
	VP1294	IHF-alpha	

	VP1283	hypothetical protein	
	VP0259	50S ribosomal protein L23	
	ureA	Urease subunit gamma	
11177.9-11216.5	VP2885	DNA-binding protein fis	Vh, Vca, Vc, Vf
	VP1294	IHF-alpha	
	VP1283	hypothetical protein	
	VP0268	50S ribosomal protein L24	
	VP0259	50S ribosomal protein L23	
	ureA	Urease subunit gamma	
11294.42	VP1294	IHF-alpha	Vp, Vh
	VP1283	hypothetical protein	
	VP0268	50S ribosomal protein L24	
11361.63	0		Vfl, Vf, Vm
12003.29	VP2986	CyaY	Vp
	VP2474	Iron-sulfur cluster insertion protein ErpA	
	VP2178	YbaB-like hypothetical protein	
	VP0262	50S ribosomal protein L22	
12089.73	VPA1222	hypothetical protein GIY-YIG catalytic domain	Vp
	VP2986	CyaY	
	VP2923	50S ribosomal protein L7/L12	
	VP2766	Met regulon regulatory protein metJ	
	VP2178	YbaB-like hypothetical protein	
	VP0262	50S ribosomal protein L22	
12563.1-12617.4	VPA1414	YpfB-like hypothetical protein	Vp, Vh, Vf, Vc
	VP2098	Phosphorelay protein LuxU	
	VP0273	50S ribosomal protein L18	
13084.43-13094.31	VPA0577	pterin-4-alpha-carbinolamine dehydratase	Vp
	VP2530	50S ribosomal protein L19	
	VP0646	hypothetical protein	
13578.65	VP3028	Protein CrcB homolog	Vp
	VP2843	Fumarate reductase subunit D	
	VP2773	30S ribosomal protein S12	
	VP2568	Holo-ACP synthase	
	VP1300	LrgA-like hypothetical protein	
	VP0267	50S ribosomal protein L14	
13751.26	VP3028	Protein CrcB homolog	Vp, Vc
	VP2843	Fumarate reductase subunit D	
	VP2773	30S ribosomal protein S12	

	VP2568	Holo-ACP synthase	
	VP0004	RNaseP protein	
13756.5-13765.3	VP0280	30S ribosomal protein S11	Vp
	VP3028	Protein CrcB homolog	
	VP2843	Fumarate reductase subunit D	
	VP2773	30S ribosomal protein S12	
	VP2568	Holo-ACP synthase	
	VP0004	RNaseP protein	
13777.16	VP2773	30S ribosomal protein S12	Vp, Vf, Vm
	VP2568	Holo-ACP synthase	
	VP0004	RNaseP protein	
	VP0280	30S ribosomal protein S11	
	VP3028	Protein CrcB homolog	
14210.99	VP0448	hypothetical protein	Vp
	VP0283	50S ribosomal protein L17	

To determine whether MALDI-TOF MS analysis could be used to identify single-gene genetic changes among strains we compared the MALDI-TOF MS fingerprints of an in-frame deletion mutant of the quorum sensing regulator *opaR* and a deletion mutant of the mismatch repair gene *mutS* to the wild-type strains (Figure 2.19). The MALDI-TOF MS spectrum of a *V. parahaemolyticus* RIMD2210633 Δ*opaR* strain had peaks of lower intensity compared to those of the wild-type strain (Figure 2.19). Overall, the number of peaks and the locations were conserved between the RIMD2210633 and Δ*opaR* spectra; however, there were several peaks that were present or absent from the RIMD or *opaR* spectra. For example, there was a peak at approximately *m/z* 6,612 and a peak at *m/z* 7,414 that were present in the RIMD spectrum and absent from the Δ*opaR* spectrum (Figure 2.19) that were predicted to correspond to a 50s ribosomal protein and a hypothetical protein (Table 2.9). There were four additional peaks of lower intensity that were present in at least two independently generated spectra of RIMD that were absent

from the *opaR* spectra (Table 2.9). The additional RIMD peaks were identified as possible hypothetical proteins, 50s and 30s ribosomal proteins, the chaperonin GroES, the integration host factor IHF-β, and a dehydratase (Table 2.9). There were two peaks present in the Δ*opaR* spectra that were repeatedly absent from the RIMD spectra at approximately *m/z* 7,154 and 9,930 (Table 2.9). The unique Δ*opaR* peaks were identified as a possible zinc-binding protein, a ribosomal protein, and the regulatory proteins Hfq and RpoZ (Table 2.9).

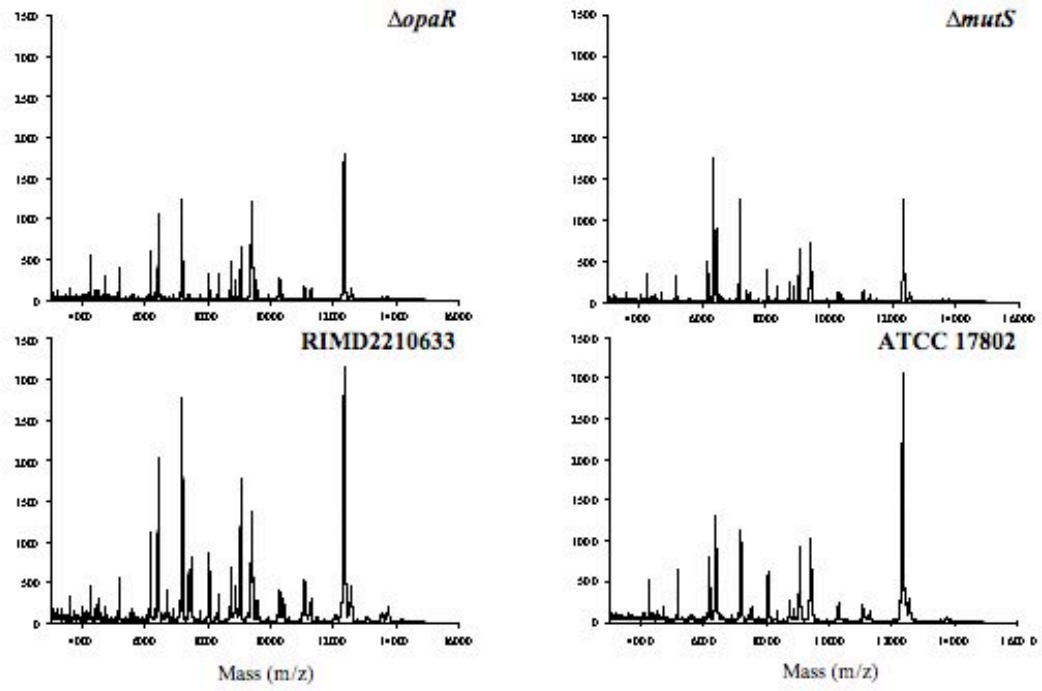


Figure 2.19. MALDI-TOF MS spectra of a *V. parahaemolyticus* RIMD2210633 Δ*opaR* strain compared to the RIMD wild-type strain, and a *V. parahaemolyticus* ATCC 17802 Δ*mutS* strain compared to the 17802 wild-type strain.

There were no visible differences in the spectra of ATCC 17802 and the Δ*mutS* mutant (Figure 2.19); however, there were several low intensity peaks detected that differed between the fingerprints of 17802 and Δ*mutS*. There was a peak present at *m/z* 6,287 in the spectrum of 17802 and absent from Δ*mutS* that corresponded to a transmembrane protein (Table 2.9). In addition, there were peaks at *m/z* 6574, 7167, 10416, and 12568 that were predicted to be 50S ribosomal proteins, CsrA, a hypothetical protein, and the phosphorelay protein LuxU that were detected for 17802 and not Δ*mutS* (Table 2.9

Table 2.9. MALDI-TOF MS peak differences of wild-type and mutant strains

Peak Mass (Da)	Predicted Genes	Predicted Proteins	Strain present
RIMD vs. *opaR*			
6612.44	*vp0275*	50S ribosomal protein L30	RIMD
7154.71	*vp2529*	zinc-binding protein	opaR
	vp0265	50S ribosomal protein L29	
7414.21	*vp2335*	hypothetical protein	RIMD
9930.69	*vp2817*	Hfq	opaR
	vp2331	50S ribosomal protein L31 type B	
	vp0160	RNAP omega subunit RpoZ	
10416.32	*vp1210*	50S ribosomal protein L25	RIMD
	vp0261	30S ribosomal protein S19	
10593.64	*vpa0321*	hypothetical protein sigma 70 region	RIMD
	vpa0286	GroES protein 2	
	vp2627	Probable Fe(2+)-trafficking protein	
	vp2619	YggU-like hypothetical protein	
	vp2029	IHF-beta	
13093.05	*vpa0577*	pterin-4-alpha-carbinolamine dehydratase	RIMD
	vp2530	50S ribosomal protein L19	
	vp0646	hypothetical protein	
14208.96	*vp0448*	hypothetical protein	RIMD
	vp0283	50S ribosomal protein L17	
17802 vs. *mutS*			
6287.00	*vp0082*	transmembrane protein	17802
6574.41	*vp0275*	50S ribosomal protein L30	17802
	vp0186	50S ribosomal protein L33	
7167.45	*vp2546*	CsrA	17802
10416.12	*vp1210*	50S ribosomal protein L25	17802
	vp0261	30S ribosomal protein S19	
12568.93	*vpa1414*	hypothetical protein, ASCH domain	17802
	vp0273	50S ribosomal protein L18	
	vp2098	Phosphorelay protein LuxU	

Discussion

An increase in the frequency of *V. parahaemolyticus*-associated infections has heightened the need for a method of rapid and reliable identification of *V. parahaemolyticus* strains (17, 48). Current methods for identification of *V. parahaemolyticus* involve biochemical tests (22), serotyping, PCR or hybridization analysis of housekeeping genes or disease-associated genes (6, 46, 55, 66, 71), MLST (14, 26), and PFGE (36, 44, 58). These methods can take a number of days in order to

accurately distinguish *V. parahaemolyticus* from other closely related *Vibrios* and require the preliminary identification of strains by growth on selective media.

In the current study, we demonstrate the development of a high-throughput whole-cell MALDI-TOF MS technique for the identification of *V. parahaemolyticus* clinical and environmental strains, which takes less than two minutes for each sample spot. We show that whole-cell MALDI-TOF MS analysis of *V. parahaemolyticus* and other *Vibrio* spp. is highly-reproducible and effective for distinguishing between related species as previously demonstrated for other bacteria (64, 74). Comparison of MALDI-TOF MS fingerprints and cluster analysis of the MALDI-TOF MS data demonstrated that whole-cell MALDI-TOF MS analysis can be used to distinguish *V. parahaemolyticus* from other *Vibrios* including the closely-related *V. harveyi* and *V. campbellii*. *V. parahaemolyticus* strains are nearly indistinguishable from *V. harveyi* and *V. campbellii* by 16S rRNA gene analysis requiring the analysis of multiple housekeeping genes (68, 69). We identified a total of 29 peaks that were specific to *V. parahaemolyticus* strains examined in this study indicating that these peaks should be further investigated for assessment as *V. parahaemolyticus* biomarkers.

In addition, we show that the *V. parahaemolyticus* clonal disease-causing strains exhibited fingerprints that had detectable differences in the number and position of peaks suggesting that MALDI-TOF MS analysis may be applied to monitor the spread and diversification of *V. parahaemolyticus* clonal pandemic strains. There was one peak detected at *m/z* 10,430 only among the *V. parahaemolyticus* O4:K12 clonal strains examined. Analysis of additional clonal O4:K12 strains would be required to determine whether this peak as well as additional peaks could serve as biomarkers for identification

of the O4:K12 strains that are frequently associated with disease outbreaks in the United States (1, 48). The more recently isolated clonal O3:K6 and O4:K12 strains exhibited different MALDI-TOF MS fingerprints compared to strains isolated in 1996 and 1997 when the clonal strains were first detected. Similarly, PFGE profiles show detectable differences among the clonal pandemic strains; however, PFGE takes many days and requires that strains must be identified as *V. parahaemolyticus* by other methods prior to analysis, which increases the time required to identify outbreak strains. In contrast, whole-cell MALDI-TOF MS analysis can be used to rapidly identify *V. parahaemolyticus* and monitor the emergence and spread of the clonal pandemic strains. Furthermore, the differences in the MALDI-TOF MS fingerprints of the clonal strains and the similarity of the co-occurring environmental strains may relate to differences in the geographic location where the strains were isolated. The analysis of additional *V. parahaemolyticus* isolated from distinct geographical regions and development of biomarker peaks would be required to support the use of MALDI-TOF MS for identification of the geographical origin of certain strain types. In this regard, MALDI-TOF MS could be a powerful tool to monitor the emergence and spread of clonal groups as they become more frequent causative agents of disease outbreaks such as the emergence of the O3:K6 clonal pandemic strain in India in 1995 (52).

MALDI-TOF MS analysis could provide greater resolution of the differences among *V. parahaemolyticus* clonal strains since changes in single genes as we have demonstrated in the current study results in a unique MALDI-TOF MS fingerprint. Therefore, fewer differences among strains can be detected that would otherwise not be resolved using PFGE or MLST. In this regard, whole-cell MALDI-TOF MS analysis

would be beneficial for monitoring recent changes among related strains such as the *V. parahaemolyticus* clonal pandemic strains. Currently, molecular techniques used for the identification of *V. parahaemolyticus* strains such as MLST and PFGE allow limited analysis of the relationships of the clonal disease-causing strains. PFGE analysis of disease-causing *V. parahaemolyticus* strains is used as a reliable tool to monitor the spread of clonal strains that have closely-related PFGE profiles (36, 58). Aside from the similarity of the clonal pandemic strains, PFGE has demonstrated that other *V. parahaemolyticus* strains have considerable diversity (36, 44, 58). MLST analysis of up to seven housekeeping genes was shown to increase the resolution for distinguishing among related strains such as the disease-causing clonal pandemic strains and resulted in the identification of two new clonal complexes (26). Although increasing the number of housekeeping genes analyzed resulted in greater resolution of the evolutionary relationships of *V. parahaemolyticus* strains, sequence analysis of a greater number of housekeeping genes also increased the time and cost associated with MLST analysis.

In the current study, we demonstrate that whole-cell MALDI-TOF MS analysis is sensitive enough to detect only a few genetic changes such as inactivation of a single regulatory gene or DNA repair pathway. Whole-cell MALDI-TOF MS analysis showed variation in the MALDI-TOF MS fingerprint of a single strain following deletion of the quorum sensing regulator *opaR* or the mismatch repair gene *mutS*. Quorum sensing has been shown to regulate type III secretion of *V. parahaemolyticus* (32) that was shown to translocate effector proteins resulting in cytotoxicity to eukaryotic cells (57). Spontaneous mutation of the quorum sensing regulator occurs naturally, resulting in the emergence of phenotypic variations such as a switch from opaque to translucent colony

morphologies (23, 45). In a previous study, we showed that deletion of the mismatch repair gene, *mutS*, resulted in a significant change in opacity that may have been caused by spontaneous mutation of *opaR* (30). The ability to monitor changes in global regulatory systems such as quorum sensing that are involved in the pathogenicity mechanism of *Vibrios* would provide additional information on the nature of disease-causing strains.

In addition, MALDI-TOF MS analysis could be used as a method of preliminary identification of *V. parahaemolyticus* strains present in a biological sample. Following further analysis and the development of biomarker peaks, this approach could be further applied to determine whether strains potentially belong to certain disease-causing groups. Comparison of the MALDI-TOF MS fingerprints of *V. parahaemolyticus* strains revealed there were detectable differences in the spectra of the clinical and environmental strains examined. Analysis of additional *V. parahaemolyticus* strains and the identification of biomarker peaks characteristic of disease-causing strains would facilitate the separation of disease-causing strains from other non-pathogenic strains present in the same sample. Recent studies have demonstrated that multiple *V. parahaemolyticus* strains can co-occur in an individual that exhibits *V. parahaemolyticus*-associated disease symptoms (7). Development of MALDI-TOF MS biomarkers for diverse *Vibrios* would facilitate the detection of mixed *Vibrio* assemblages that could be applied to determine how many different *Vibrio* spp. are present in a sample. Recently, whole-cell MALDI-TOF MS analysis was used to detect *Erwinia* spp. that were cultured from different biological samples (64). A similar approach could be used to identify disease-causing *Vibrios* by inoculating and growing a mixed sample under standardized conditions followed by

MALDI analysis. The presence of certain species-specific biomarker peaks or peaks characteristic of particular groups of disease-causing strains would allow rapid detection of the presence of a particular *Vibrio* pathogen.

In this study, we report the first use of whole-cell MALDI-TOF MS analysis as a powerful tool for identification of *V. parahaemolyticus* strains. Based on the findings of our study, the rapid and reliable generation of whole-cell MALDI-TOF MS fingerprints would be an important tool for the initial identification of *V. parahaemolyticus* and other *Vibrio* spp. In addition, we identified potential biomarker peaks that could be further developed for detection of *V. parahaemolyticus* and analysis of differences among disease-causing strains. Further application of this method would involve the construction of a database of whole-cell MALDI-TOF MS fingerprints from *V. parahaemolyticus* and related disease-causing *Vibrios* that could be referenced for identification of the causative agents of disease outbreaks. Research is ongoing to determine whether the whole-cell MALDI-TOF MS approach described in this study may be used to detect multiple disease-causing *Vibrios* present in food or environmental samples.

Acknowledgements

I would like to thank R. Martinez and Y. Chen for their work developing techniques and generating data analyzed in this study. Also, I would like to thank N. Garrett for the serotyping performed and P. Lafon for PFGE analysis.

References

1. **(CDC), C. f. D. C. a. P.** 2006. *Vibrio parahaemolyticus* infections associated with consumption of raw shellfish--three states, 2006. MMWR Morb. Mortal. Wkly. Rep. **55:**854-856.
2. **Ansaruzzaman, M., M. Lucas, J. L. Deen, N. A. Bhuiyan, X. Y. Wang, A. Safa, M. Sultana, A. Chowdhury, G. B. Nair, D. A. Sack, L. von Seidlein, M. K. Puri, M. Ali, C. L. Chaignat, J. D. Clemens, and A. Barreto.** 2005. Pandemic serovars (O3:K6 and O4:K68) of *Vibrio parahaemolyticus* associated with diarrhea in Mozambique: spread of the pandemic into the African continent. J. Clin. Microbiol. **43:**2559-2562.
3. **Baba, K., H. Shirai, A. Terai, K. Kumagai, Y. Takeda, and M. Nishibuchi.** 1991. Similarity of the *tdh* gene-bearing plasmids of *Vibrio cholerae* non-O1 and *Vibrio parahaemolyticus*. Microb. Pathog. **10:**61-70.
4. **Bagwell, C. E., Y. M. Piceno, A. Ashburne-Lucas, and C. R. Lovell.** 1998. Physiological diversity of the rhizosphere diazotroph assemblages of selected salt marsh grasses. Appl. Environ. Microbiol. **64:**4276-4282.
5. **Barbuddhe, S. B., T. Maier, G. Schwarz, M. Kostrzewa, H. Hof, E. Domann, T. Chakraborty, and T. Hain.** 2008. Rapid identification and typing of *Listeria* species by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. **74:**5402-5407.
6. **Bej, A. K., D. P. Patterson, C. W. Brasher, M. C. Vickery, D. D. Jones, and C. A. Kaysner.** 1999. Detection of total and hemolysin-producing *Vibrio parahaemolyticus* in shellfish using multiplex PCR amplification of *tl, tdh,* and *trh.* J. Microbiol. Methods **36:**215-225.
7. **Bhoopong, P., P. Palittapongarnpim, R. Pomwised, A. Kiatkittipong, M. Kamruzzaman, Y. Nakaguchi, M. Nishibuchi, M. Ishibashi, and V. Vuddhakul.** 2007. Variability of properties of *Vibrio parahaemolyticus* strains isolated from individual patients. J. Clin. Microbiol. **45:**1544-1550.
8. **Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453-74.
9. **Boyd, E. F., A. L. V. Cohen, L. M. Naughton, D. W. Ussery, T. T. Binnewies, O. C. Stine, and M. A. Parent.** 2008. Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. BMC Microbiol. **8:**110-124.
10. **Buck, J. D., R. S. Wells, H. L. Rhinehart, and L. J. Hansen.** 2006. Aerobic microorganisms associated with free-ranging bottlenose dolphins in coastal Gulf of Mexico and Atlantic Ocean waters. J. Wild. Dis. **42:**536-544.
11. **Burdette, D. L., M. L. Yarbrough, A. Orvedahl, C. J. Gilpin, and K. Orth.** 2008. *Vibrio parahaemolyticus* orchestrates a multifaceted host cell infection by induction of autophagy, cell rounding, and then cell lysis. Proc. Nat. Acad. of Sci. USA **105:**12497-12502.
12. **Cabanillas-Beltrán, H., E. Llausás-Magaña, R. Romero, A. Espinoza, A. García-Gasca, M. Nishibuchi, M. Ishibashi, and B. Gomez-Gil.** 2006.

Outbreak of gastroenteritis caused by the pandemic *Vibrio parahaemolyticus* O3:K6 in Mexico. FEMS Microbiol. Lett. **265:**76-80.

13. **Chang, B., H. Miyamoto, H. Taniguchi, and S. Yoshida.** 2002. Isolation and genetic characterization of a novel filamentous bacteriophage, a deleted form of phage f237, from a pandemic *Vibrio parahaemolyticus* O4:K68 strain. Microbiol. Immunol. **46:**565-569.
14. **Chowdhury, N. R., O. C. Stine, J. Glenn Morris, and G. B. Nair.** 2004. Assessment of evolution of pandemic *Vibrio parahaemolyticus* by multilocus sequence typing. J. Bacteriol. **42:**1280-1282.
15. **Chun, D., J. K. Chung, R. Tak, and S. Y. Seol.** 1975. Nature of the Kanagawa phenomenon of *Vibrio parahaemolyticus*. Infect. Immun. **12:**81-87.
16. **Criminger, J. D., T. H. Hazen, P. A. Sobecky, and C. R. Lovell.** 2007. Nitrogen fixation by *Vibrio parahaemolyticus* and its implications for a new ecological niche. Appl. Environ. Microbiol. **73:**5959-5961.
17. **Daniels, N. A., L. MacKinnon, R. Bishop, S. Altekruse, B. Ray, R. M. Hammond, S. Thompson, S. Wilson, N. H. Bean, P. M. Griffin, and L. Slutsker.** 2000. *Vibrio parahaemolyticus* infections in the United States, 1973-1998. J. Infect. Dis. **181:**1661-1666.
18. **DePaola, A., J. Ulaszek, C. A. Kaysner, B. J. Tenge, J. L. Nordstrom, J. Wells, N. Puhr, and S. M. Gendel.** 2003. Molecular, serological, and virulence characteristics of *Vibrio parahaemolyticus* isolated from environmental, food, and clinical sources in North America and Asia. Appl. Environ. Microbiol. **69:**3999-4005.
19. **Di Lorenzo, M., M. Stork, M. E. Tolmasky, L. A. Actis, D. Farrell, T. J. Welch, L. M. Crosa, A. M. Wertheimer, Q. Chen, P. Salinas, L. Waldbeser, and J. H. Crosa.** 2003. Complete sequence of virulence plasmid pJM1 from the marine fish pathogen *Vibrio anguillarum* strain 775. J. Bacteriol. **185:**5822-5830.
20. **Dieckmann, R., R. Helmuth, M. Erhard, and B. Malorny.** 2008. Rapid classification and identification of Salmonellae at the species and subspecies levels by whole-cell matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. **74:**7767-7778.
21. **Dziejman, M., D. Serruto, V. C. Tam, D. Sturtevant, P. Diraphat, S. M. Faruque, M. H. Rahman, J. F. Heidelberg, J. Decker, L. Li, K. T. Montgomery, G. Grills, R. Kucherlapati, and J. J. Mekalanos.** 2005. Genomic characterization of non-O1, non-O139 *Vibrio cholerae* reveals genes for a type III secretion system. Proc. Natl. Acad. Sci. USA **102:**3465-3470.
22. **Elliot, E. L., C. A. Kaysner, L. Jackson, and M. L. Tamplin.** 1998. *Vibrio cholerae, V. parahaemolyticus, V. vulnificus,* and other *Vibrio* spp., p. 9.01-9.27. *In* U. S. F. a. D. A. B. A. Manual (ed.). A.O.A.C. International, Gaithersburg, MD.
23. **Enos-Berlage, J. L., Z. T. Guvener, C. E. Keenan, and L. L. McCarter.** 2005. Genetic determinants of biofilm development of opaque and translucent *Vibrio parahaemolyticus*. Mol. Microbiol. **55:**1160-1182.
24. **Fujino, T., T. Miwatani, J. Yasuda, M. Kondo, Y. Takeda, Y. Akita, K. Kotera, M. Okada, H. Nishimune, Y. Shimizu, T. Tamura, and Y. Tamura.** 1965. Taxonomic studies on the bacterial strains isolated from cases of "shirasu"

food-poisoning (*Pasteurella parahaemolytica*) and related microorganisms. Biken J. **8:**63-71.

25. **Gasteiger, E., C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch.** 2005. Protein identification and analysis tools on the ExPASy server, p. 571-607. *In* J. M. Walker (ed.), The proteomics protocols handbook. Humana Press.
26. **González-Escalona, N., J. Martinez-Urtaza, J. Romero, R. T. Espejo, L. Jaykus, and A. DePaola.** 2008. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. J. Bacteriol. **190:**2831-2840.
27. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and anlysis program for Windows 95/98/NT. Nucleic Acids Symp Ser **41:**95-98.
28. **Han, H., H. C. Wong, B. Kan, Z. Guo, X. Zeng, S. Yin, X. Liu, D. Zhou, and R. Yang.** 2009. Genome plasticity of *Vibrio parahaemolyticus:* microevolution of the 'pandemic group'. Bmc Genomics.
29. **Hazen, T. H., D. J. Silberger, D. Wu, Garrett, N., Bopp, C. A., J. A. Eisen, and P. A. Sobecky.** In Submission. Diversity of mobile genetic elements isolated from *Vibrio parahaemolyticus* clinical and environmental strains and the identification of a plasmid-encoded bacteriocin.
30. **Hazen, T. H., K. D. Kennedy, S. Chen, S. V. Yi, and P. A. Sobecky.** 2009. Inactivation of mismatch repair increases the diversity of *Vibrio parahaemolyticus*. Env. Microbiol.**:**10.1111/j.1462-2920.2008.01853.x
31. **Hazen, T. H., D. Wu, J. A. Eisen, and P. A. Sobecky.** 2007. Sequence characterization and comparative analysis of three plasmids isolated from environmental *Vibrio* spp. Appl. Environ. Microbiol. **73:**7703-7710.
32. **Henke, J. M., and B. L. Bassler.** 2004. Quorum sensing regulates type III secretion in *Vibrio harveyi* and *Vibrio parahaemolyticus*. J. Bacteriol. **186:**3794-3805.
33. **Hunt, D. E., L. A. David, D. Gevers, S. P. Preheim, E. J. Alm, and M. F. Polz.** 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science **320:**1081-1085.
34. **Hurley, C. C., A. M. Quirke, F. J. Reen, and E. F. Boyd.** 2006. Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates. BMC Genom. **7:**104.
35. **Izutsu, K., K. Kurokawa, K. Tashiro, S. Kuhara, T. Hayashi, T. Honda, and T. Iida.** 2008. Comparative genomic analysis using microarray demonstrates a strong correlation between the presence of the 80-kilobase pathogenicity island and pathogenicity in Kanagawa phenomenon-positive *Vibrio parahaemolyticus* strains. Infect. Immun. **76:**1016-1023.
36. **Kam, K. M., C. K. Y. Luey, M. B. Parsons, K. L. F. Cooper, G. B. Nair, M. Alam, M. A. Islam, D. T. L. Cheung, Y. W. Chu, T. Ramamurthy, G. P. Pazhani, S. K. Bhattacharya, H. Watanabe, J. Terajima, E. Arakawa, O.-A. Ratchtrachenchai, S. Huttayananont, E. M. Ribot, P. Gerner-Smidt, and B. Swaminathan.** 2008. Evaluation and validation of a PulseNet standardized pulsed-field gel electrophoresis protocol for subtyping *Vibrio parahaemolyticus:* an international multicenter collaborative study. J. Clin. Microbiol. **46:**2766-2773.

37. **Kamruzzaman, M., and M. Nishibuchi.** 2008. Detection and characterization of a functional insertion sequence, ISVpa2, in *Vibrio parahaemolyticus*. Gene **409:**92-99.
38. **Kim, Y. K., and L. L. MCCarter.** 2007. ScrG, a GGDEF-EAL protein, participates in regulating swarming and sticking in *Vibrio parahaemolyticus*. J. Bacteriol. **189:**4094-4107.
39. **Kobayashi, T., S. Enomoto, R. Sakazaki, and S. Kuwahara.** 1963. A new selective isolation medium for the *Vibrio* group; on a modified Nakanishi's medium (TCBS agar medium). Nippon Saikingaku Zasshi **18:**387-92.
40. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinf. **5:**150-163.
41. **Lay, J. O.** 2001. MALDI-TOF mass spectrometry of bacteria. Mass Spectrom. Rev. **20:**172-194.
42. **Lee, C. T., C. Amaro, K. M. Wu, E. Valiente, Y. F. Change, S. F. Tsai, C. H. Chang, and L. I. Hor.** 2008. A common virulence plasmid in biotype 2 *Vibrio vulnificus* and its dissemination aided by a conjugal plasmid. J. Bacteriol. **190:**1638-48.
43. **Makino, K., K. Oshima, K. Kurokawa, K. Yokoyama, T. Uda, K. Tagomori, Y. Iijima, M. Najima, M. Nakano, A. Yamashita, Y. Kubota, S. Kimura, T. Yasunaga, T. Honda, H. Shinagawa, M. Hattori, and T. Iida.** 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. Lancet **361:**743-749.
44. **Martinez-Urtaza, J., A. Lozano-Leon, A. DePaola, M. Ishibashi, K. Shimada, M. Nishibuchi, and E. Liebana.** 2004. Characterization of pathogenic *Vibrio parahaemolyticus* isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates. J. Clin. Microbiol. **42:**4672-8.
45. **McCarter, L. L.** 1998. OpaR, a homolog of *Vibrio harveyi* LuxR, controls opacity of *Vibrio parahaemolyticus*. J. Bacteriol. **180:**3166-3173.
46. **Meador, C. E., M. A. Parsons, C. A. Bopp, P. Gerner-Smidt, J. A. Painter, and G. J. Vora.** 2007. Virulence gene- and pandemic group-specific marker profiling of clinical *Vibrio parahaemolyticus* isolates. J. Clin. Microbiol. **45:**1133-1139.
47. **Myers, M. L., G. Panicker, and A. K. Bej.** 2003. PCR detection of a newly emerged pandemic *Vibrio parahaemolyticus* O3:K6 pathogen in pure cultures and seeded waters from the Gulf of Mexico. Appl. Environ. Microbiol. **69:**2194-2200.
48. **Nair, G. B., T. Ramamurthy, S. K. Bhattacharya, B. Dutta, Y. Takeda, and D. A. Sack.** 2007. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. Clin. Microbiol. Rev. **20:**39-48.
49. **Nasu, H., T. Iida, T. Sugahara, Y. Yamaichi, K. S. Park, K. Yokoyama, K. Makino, H. Shinagawa, and T. Honda.** 2000. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. J. Clin. Microbiol. **38:**2156-2161.
50. **Nei, M.** 1987. Molecular Evolutionary Genetics, vol. Columbia University Press, New York.

51. **Okada, N., T. Iida, K. Park, N. Goto, T. Yasunaga, H. Hiyoshi, S. Matsuda, T. Kodama, and T. Honda.** 2009. Identification and characterization of a novel type III secretion system in *trh*-positive *Vibrio parahaemolyticus* strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level. Infect. Immun. **77:**904-913.
52. **Okuda, J., M. Ishibashi, E. Hayakawa, T. Nishino, Y. Takeda, A. K. Mukhopadhyay, S. Garg, S. K. Bhattacharya, G. B. Nair, and M. Nishibuchi.** 1997. Emergence of a unique O3:K6 clone of *Vibrio parahaemolyticus* in Calcutta, India, and isolation of strains from the same clonal group from southeast Asian travelers arriving in Japan. J. Clin. Microbiol. **35:**3150-3155.
53. **Okura, M., R. Osawa, E. Arakawa, J. Terajima, and H. Watanabe.** 2005. Identification of *Vibrio parahaemolyticus* pandemic group-specific DNA sequence by genomic subtraction. J. Clin. Microbiol. **43:**3533-3536.
54. **Ottaviani, D., F. Leoni, E. Rocchegiani, S. Santarelli, C. Canonico, L. Masini, V. Ditrani, and A. Carraturo.** 2008. First clinical report of pandemic *Vibrio parahaemolyticus* O3:K6 infection in Italy. J. Clin. Microbiol. **46:**2144-2145.
55. **Panicker, G., D. R. Call, M. J. Krug, and A. K. Bej.** 2004. Detection of pathogenic *Vibrio* spp. in shellfish by using multiplex PCR and DNA microarrays. Appl. Environ. Microbiol. **70:**7436-7444.
56. **Park, K. S., M. Arita, T. Iida, and T. Honda.** 2005. *vpaH,* a gene encoding a novel histone-like nucleoid structure-like protein that was possibly horizontally acquired, regulates the biogenesis of lateral flagella in *trh*-positive *Vibrio parahaemolyticus* TH3996. Infect. Immun. **73:**5754-5761.
57. **Park, K. S., T. Ono, M. Rokuda, M. H. Jang, K. Okada, T. Iida, and T. Honda.** 2004. Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. Infect. Immun. **72:**6659-6665.
58. **Parsons, M. B., K. L. Cooper, K. A. Kubota, N. Puhr, S. Simington, P. S. Calimlim, D. Schoonmaker-Bopp, C. Bopp, B. Swaminathan, P. Gerner-Smidt, and E. M. Ribot.** 2007. PulseNet USA standardized pulsed-field gel electrophoresis protocol for subtyping of *Vibrio parahaemolyticus*. Foodborne Pathog. Dis. **4:**285-92.
59. **Pignone, M., K. M. Greth, J. Cooper, D. Emerson, and J. Tang.** 2006. Identification of mycobacteria by matrix-assisted laser desorption ionization-time-of-flight mass spectrometry. J. Clin. Microbiol. **44:**1963-1970.
60. **Quilici, M. L., A. Robert-Pillot, J. Picart, and J. M. Fournier.** 2005. Pandemic *Vibrio parahaemolyticus* O3:K6 spread, France. Emerg. Infec. Dis. **11:**1148-1149.
61. **Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas.** 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:**2496-7.
62. **Ruby, E. G., M. Urbanowski, J. Campbell, A. Dunn, M. Faini, R. Gunsalus, P. Lostroh, C. Lupp, J. McCann, D. Millikan, A. Schaefer, E. Stabb, A. Stevens, K. Visick, C. Whistler, and E. P. Greenberg.** 2005. Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners. Proc. Natl. Acad. Sci. USA **102:**3004-3009.

63. **Sakurai, J., A. Matsuzaki, Y. Takeda, and T. Miwatani.** 1974. Existence of two distinct hemolysins in *Vibrio parahaemolyticus*. Infect. Immun. **9:**777-780.
64. **Sauer, S., A. Freiwald, T. Maier, M. Kube, R. Reinhardt, M. Kostrzewa, and K. Geider.** 2008. Classification and identification of bacteria by mass spectrometry and computational analysis. Plos One **3:**e2843.
65. **Tamano, K., S. I. Aizawa, E. Katayama, T. Nonaka, S. Imajoh-Ohmi, A. Kuwae, S. Nagai, and Sasakawa.** 2000. Supramolecular structure of the *Shigella* type III secretion machinery: the needle part is changeable in length and essential for delivery of effectors. EMBO J. **15:**3876-3887.
66. **Tarr, C. L., J. S. Patel, N. D. Puhr, E. G. Sowers, C. A. Bopp, and N. A. Strockbine.** 2007. Identification of *Vibrio* isolates by a multiplex PCR assay and *rpoB* sequence determination. J. Clin. Microbiol. **45:**134-40.
67. **Terai, A., K. Baba, H. Shirai, O. Yoshida, Y. Takeda, and M. Nishibuchi.** 1991. Evidence for insertion sequence-mediated spread of the thermostable direct hemolysin gene among Vibrio species. J. Bacteriol. **173:**5036-5046.
68. **Thompson, F. L., D. Gevers, C. C. Thompson, P. Dawyndt, S. Naser, B. Hoste, C. B. Munn, and J. Swings.** 2005. Phylogeny and molecular identification of *Vibrios* on the basis of multilocus sequence analysis. Appl. Environ. Microbiol. **71:**5107-5115.
69. **Thompson, F. L., B. Gomez-Gil, A. T. R. Vasconcelos, and T. Sawabe.** 2007. Multilocus sequence analysis reveals that *Vibrio harveyi* and *V. campbellii* form distinct species. Appl. Environ. Microbiol. **epub ahead of print**.
70. **Thompson, F. L., T. Iida, and J. Swings.** 2004. Biodiversity of *Vibrios*. Microbiol. Mol. Biol. Rev. **68:**403-431.
71. **Vora, G. J., C. E. Meador, M. M. Bird, C. A. Bopp, J. D. Andreadis, and D. A. Stenger.** 2005. Microarray-based detection of genetic heterogeneity, antimicrobial resistance, and the viable but nonculturable state in human pathogenic *Vibrio* spp. Proc. Natl. Acad. Sci. USA **102:**19109-19114.
72. **Vuddhakul, V., A. Chowdhury, V. Laohaprertthisan, P. Pungrasamee, N. Patararungrong, P. Thianmontri, M. Ishibashi, C. Matsumoto, and M. Nishibuchi.** 2000. Isolation of a pandemic O3:K6 clone of a *Vibrio parahaemolyticus* strain from environmental and clinical sources in Thailand. Appl. Environ. Microbiol. **66:**2685-2689.
73. **Watterson, G. A.** 1975. On the number of segregation sites. Theoret. Popul. Biol. **7:**256-276.
74. **Williamson, Y. M., H. Moura, A. R. Woolfitt, J. L. Pirkle, J. R. Barr, M. D. G. Carvalho, E. P. Ades, G. M. Carlone, and J. S. Sampson.** 2008. Differentiation of *Streptococcus pneumoniae* conjunctivitis outbreak isolates by matrix-assisted laser desorption ionization-time of flight mass spectrometry. Appl. Environ. Microbiol. **74:**5891-5897.
75. **Yeung, P. S. M., M. C. Hayes, A. DePaola, C. A. Kaysner, L. Kornstein, and K. J. Boor.** 2002. Comparative phenotypic, molecular, and virulence characterization of *Vibrio parahaemolyticus* O3:K6 isolates. Appl. Environ. Microbiol. **68:**2901-2909.

CHAPTER 3

DIVERSITY OF *VIBRIO* PLASMIDS AND THEIR ROLE IN THE EVOLUTION OF *VIBRIO PARAHAEMOLYTICUS* CLINICAL AND ENVIRONMENTAL STRAINS

This chapter discusses the role of extrachromosomal genetic elements in the evolution of *V. parahaemolyticus.* I examined the diversity of plasmids isolated from environmental *Vibrios* and *V. parahaemolyticus* clinical and environmental strains to determine the role of plasmids in both intraspecific gene transfer among *V. parahaemolyticus* strains and interspecific gene transfer between *V. parahaemolyticus* and related *Vibrios.* The main objective of this chapter was to determine the contribution of plasmids to HGT that is driving the evolution of *V. parahaemolyticus* strains. I hypothesized that plasmids of *V. parahaemolyticus* strains and related *Vibrios* transfer genes that confer novel adaptive and pathogenic phenotypes.

The objectives of part 1 were to 1) examine the sequence diversity of plasmids of *Vibrios* isolated from the environment and 2) determine whether these plasmids are involved in frequent horizontal gene transfer (HGT) among co-occurring *Vibrios*. Three plasmids were characterized that encoded phage-like genes suggesting a role of *Vibrio* plasmids in acquiring new genetic material by recombination with phage. In addition, the plasmids encoded housekeeping genes including *recA* showing that *Vibrio* plasmids are involved in the evolution of conserved genes and that *recA* may not be a suitable marker for host strain identification.

The second part of this chapter examines the diversity of plasmids among *V. parahaemolyticus* strains from clinical and environmental sources and determines the relatedness of these plasmids to plasmids isolated from closely related *Vibrios*. I examined the diversity of plasmids among *V. parahaemolyticus* strains of different origin and similarity of plasmids among related *Vibrios* to determine whether plasmids are involved in intraspecific and interspecific gene exchange contributing to the evolution of *V. parahaemolyticus*. The objectives of part 2 were to 1) examine the diversity and distribution of related plasmids among *V. parahaemolyticus* and co-occurring *V. harveyi* and *V. campbellii* strains and 2) to determine whether *V. parahaemolyticus* plasmids are involved in HGT that contributes to the evolution of pathogenicity or environmental survival phenotypes of *V. parahaemolyticus* clinical and environmental strains. Sequence characterization of a 28-kb plasmid isolated from a *V. parahaemolyticus* environmental strain indicated the plasmid had a conserved replication protein-encoding gene that was present in other strains. In addition, the *V. parahaemolyticus* plasmid encoded a putative bacteriocin from the VPaI-6 genomic island of pandemic strains. The results of this section indicated a role of *V. parahaemolyticus* plasmids in both intraspecific and interspecific gene exchange among closely related environmental *Vibrios* and the contribution of these plasmids to the emergence of the pandemic disease-causing strains.

Part 1. Sequence Characterization and Comparative Analysis of Three Plasmids Isolated From Environmental *Vibrio* spp.

Abstract

The horizontal transfer of genes by mobile genetic elements such as plasmids and phage can accelerate genome diversification of *Vibrio* spp. affecting their physiology, pathogenicity, and ecological character. In this study, sequence analysis of three plasmids from *Vibrio* spp. previously isolated from salt marsh sediment revealed the remarkable diversity of these elements. Plasmids p0908 (81.4-kb), p23023 (52.5-kb), and p09022 (31.0-kb) had 99, 64, and 32 predicted protein coding sequences, and G+C contents of 49.2%, 44.7%, and 42.4%; respectively. A phylogenetic tree based on concatenation of the host 16S rRNA and *rpoA* nucleotide sequences indicated p23023 and p09022 were isolated from strains most closely related to *V. mediterranei* and *V. campbellii,* respectively, while the host of p0908 forms a clade with *V. fluvialis* and *V. furnissii*. Many predicted proteins had amino acid identity to proteins of previously characterized phage and plasmids (24-94%). Predicted proteins with similarity to chromosomally encoded proteins included RecA, a nucleoid associated protein (NdpA), a type IV helicase (UvrD), and multiple hypothetical proteins. Plasmid p0908 had striking similarity to Enterobacteria phage P1 sharing genetic organization and amino acid identity for 23 predicted proteins. This study provides evidence of genetic exchange between *Vibrio* plasmids, phage, and chromosomes among diverse *Vibrio* spp.

Introduction

The *Vibrionaceae* are gram negative γ-proteobacteria that occur in temperate to tropical, coastal, and estuarine marine systems (109). *Vibrio* spp. occupy a diverse range of ecological niches including sediments, the water column, and in association with organisms either as symbionts (90) or pathogens (46, 61). Phage contribute to *Vibrio* evolution and ecology by regulating host abundance (51) and transferring virulence genes such as the cholera toxin encoded by *ctxAB* of the CTXϕ phage of *V. cholerae* (112). Plasmids such as pJM1 of *V. anguillarum* (33) have also been shown to play a role in *Vibrio* pathogenicity. In recent years, sequencing has revealed the vast diversity of phage genomes (15) and their globally significant contributions to horizontal gene transfer within marine environments (59). In contrast to the demonstrated genetic diversity of vibriophage (26, 114), much less is known of *Vibrio* plasmid diversity and the role of plasmids in gene transfer. A few studies have reported the occurrence of plasmids among *Vibrio* populations (31, 33, 34, 83, 111) and several have reported complete sequences of *Vibrio* plasmids associated with pathogenic *Vibrios*; however, the distribution and sequence diversity of *Vibrio* plasmids has not been studied as extensively as vibriophage.

As of September 2007 there are 16 plasmid and 20 phage sequences in Genbank that were isolated from *Vibrios* (20-22, 33, 34, 37, 42, 47, 50, 53, 67, 72, 77, 84, 86, 89, 90, 93, 115). These sequences are biased toward small elements (9 plasmids < 8-kb, 10 phage < 9-kb) and are primarily associated with well-characterized human and fish pathogens. Among these are plasmids isolated from *V. anguillarum* (33, 115), *V. cholerae* (86, 89), *V. vulnificus* (22), *V. parahaemolyticus* (72), and *V. salmonicida*. The lack of plasmid sequence data, particularly of plasmids from *Vibrio* hosts isolated from

coastal water and sediment, limits our understanding of *Vibrio* plasmid evolution and diversity.

In the present study we provide a comparative assessment of plasmids with diverse sizes and gene contents isolated from *Vibrios*. Similarities of replication initiation and hypothetical proteins revealed relatedness of plasmids from *Vibrios* occupying diverse niches. In addition, these elements contained numerous phage-like proteins including proteins with considerable similarity and conserved gene order to Enterobacteria phage P1. To our knowledge, this is the first characterization of an extrachromosomal element with P1-like phage sequences isolated from a marine bacterium. A previous study identified two P1-like genes as part of a marine viral metagenome (15); and a search of the P1 nucleotide sequences to a metagenome database yielded a few hits to marine microbial metagenomes indicating additional MGEs however, no additional P1 genes or nearly complete P1 genomes have been characterized from the marine environment.

Materials and Methods

Bacterial strains, media, and plasmid isolation. *Vibrio* spp. 0908, 23023, and 09022 were isolated from salt marsh sediment of Charleston, South Carolina in December 1998 (28). DNA for sequencing was obtained by purification of supercoiled plasmid DNA by cesium chloride density gradient centrifugation as previously described (96).

Plasmid sequencing and sequence analysis. Plasmids were sequenced using whole genome shotgun sequencing and finishing methods (46). Initial open reading frame (ORF) designation, and annotation of select ORFs was done using an automated

annotation system (46). Protein coding sequences (CDSs) were confirmed by independent analysis using GeneMark software (11). Putative similarity to known proteins was determined by amino acid sequence comparison and identification of common motif and domain structure using a combination of PSI-BLAST (5) from the National Center for Biotechnology Information (NCBI), SMART (92), COG (102, 103), and Pfam (9) web-based software. PSI-BLAST analysis was performed with the default threshold E-value of 0.005 and a maximum threshold of 1.0 over one to two iterations. ClustalW was used to generate all alignments (108).

Phylogenetic analyses and sequence alignments. Host strains were identified by a concatenated phylogenetic analysis of 16S rRNA and *rpoA* nucleotide sequences as previously described (29). The neighbor-joining tree was generated using MEGA with the Jukes-Cantor (52) distance estimation model with 1,000 replications for the nucleotide concatenation or the Poisson correction for the amino acid RecA tree (73). Percent identities of the nucleotide sequences to the most related organism were determined using BLASTN (5) and BLAST 2 sequences (104). Sequencing was performed by the University of Nevada, Reno Genomics Center and the Core Genomics Facility at the Georgia Institute of Technology.

Identification of phage-like proteins. Prophage Finder (13) was used with BLAST analysis (5, 91) to a phage sequence database to identify prophage and proteins with similarity to phage-associated proteins for all sequenced *Vibrio* plasmids available in Genbank as of July 2007. An E-value of 0.001 with 10 hits/prophage and hit spacing of 3500 were used as parameters for all plasmids examined.

Accession Numbers. The plasmid sequences have been submitted to the Genbank database under accession numbers CP000755-CP000757. All additional sequences have been submitted to the Genbank database under accession numbers EU022567-EU022572.

Results and Discussion

Host phylogeny and plasmid features. In this study we examined the sequence diversity of plasmids previously isolated from three *Vibrio* hosts (28). A concatenation of 16S rRNA sequences and *rpoA* nucleotide sequences was used for greater resolution of related *Vibrio* spp. (106). The 16S rRNA and *rpoA* nucleotide sequences of *Vibrio* sp. 0908, 23023, and 09022 were 98 and 97%, 98 and 99%, and 99 and 98% identical to those of *V. fluvialis, V. mediterranei,* and *V. campbellii;* respectively. Phylogenetic analysis of concatenated 16S rRNA and *rpoA* nucleotide sequences of *Vibrio* sp. 23023 and 09022 indicated they were most related to *V. mediterranei,* and *V. campbellii*; respectively (Figure 3.1). *Vibrio* sp. 0908 forms a clade with the closely-related *V. furnissii* and *V. fluvialis* group (16). To date, the only report of MGEs associated with any of these *Vibrio* species is an SXT-like element of *V. fluvialis* with similarity to the multiple-antibiotic resistance element SXT previously characterized from *V. cholerae* (4). This previous study indicates there may be transfer of MGEs among well-characterized pathogens such as *V. cholerae* and emerging marine pathogens such as *V. fluvialis* (16, 57, 100).

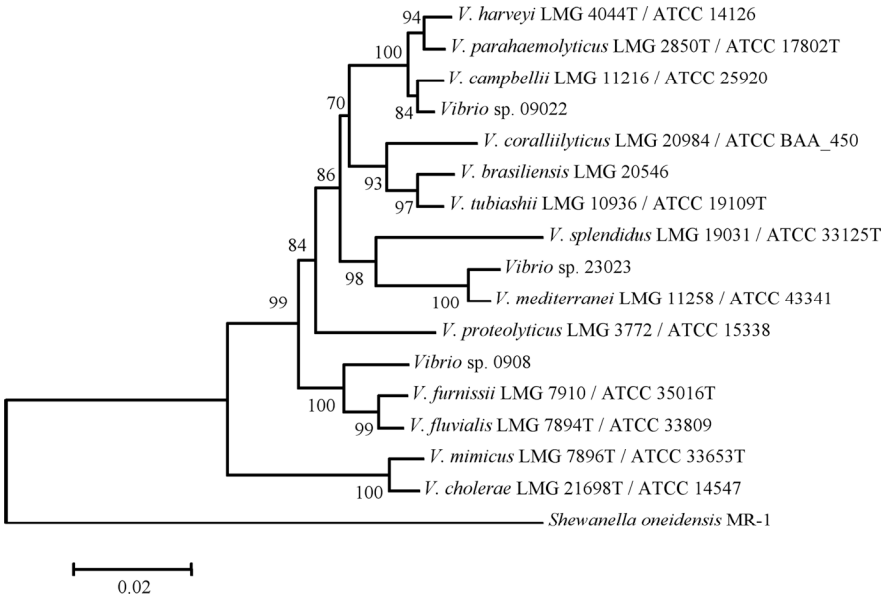


Figure 3.1. A concatenation of 16S rRNA and *rpoA* nucleotide sequences of the plasmid hosts, *Vibrio* sp. 0908, 23023, and 09022 was used to determine relatedness of the hosts to other *Vibrio* spp. as examined in a previous study (106). The neighbor-joining method with the Jukes-Cantor (52) model of distance estimation was used to generate the tree with a concatenation of 16S rRNA (1452 nt) and *rpoA* (772 nt) nucleotide sequences. Bootstrap values represent 1,000 replications and only those with values ≥ 50 are shown.

The nucleotide sequences of the *Vibrio* plasmids p0908, p23023, and p09022 were 81,413-bp, 52,527-bp, and 31,036-bp in length with an overall G+C content of 49.2%, 44.7%, and 42.4%; respectively (Table 3.1). With the exception of p0908, the G+C contents of the plasmids were within the range of percentages reported for *Vibrio* genomes (38-47%) (22, 46, 61, 90). The plasmids p0908, p23023, and p09022 encoded 99, 64, and 32 predicted CDSs, respectively (Table 3.1). The predicted proteins were assigned primarily to the following functional categories: replication, stable maintenance, partitioning, and recombination. Additional predicted proteins identified on one or more of the plasmids may be involved in mobilization, restriction modification, or transcriptional regulation (Tables 3.2-3.4). The only genes common to at least two of the three plasmids were the putative replication initiation and partitioning proteins. The predicted replication initiation protein of p09022 encoded by CDS19 was 94% identical to the replication initiation protein of plasmid pKA1 from *V. cholerae* and 39% identical to the replication initiation protein of p0908. The predicted protein of p23023 most closely resembling a replication initiation protein was that encoded by CDS11 although it had little similarity to predicted replication proteins from characterized *Vibrio* or other marine plasmids.

Table 3.1. Sequence features of the environmental *Vibrio* plasmid sequences examined in this study

Sequence features	p0908	p23023	p09022
Size	81,413 bp	52,527 bp	31,036 bp
G+C content	49.2%	44.7%	42.4%
No. CDSs	99	64	32
Max. protein size	2350 aa	862 aa	666 aa
Average protein size	247 aa	232 aa	289 aa
No. known proteins	34 (34%)	8 (12.3%)	7 (23.3%)
No. conserved hypothetical proteins	1 (1%)	7 (10.8%)	1 (3.3%)
No. hypothetical proteins	22 (22%)	21 (32.3%)	11 (36.6%)

Plasmids encoding putative proteins for self-mobilization, such as p23023, may be frequently transferred between *Vibrio* hosts. In contrast, plasmids such as p0908 and p09022 without identifiable proteins aiding transfer may rely on transmission by phage or other mechanisms. *V. cholerae* was recently shown to naturally transform 22-kb segments of genomic DNA suggesting mechanisms of DNA uptake may facilitate incorporation of large DNA molecules (68). Additional studies would be required to determine mechanisms promoting transmission of the plasmids described in this study.

The significant amino acid identity (≥ 83%; Table 3.2) and conserved gene order of six predicted proteins encoded of p09022 compared to those of *V. cholerae* plasmid pKA1 suggests *Vibrio* plasmids from diverse hosts may undergo frequent gene exchange.

Alternately, this may indicate a common *rep* family exists among diverse *Vibrio* hosts as the conserved genes included 3 proteins likely to be involved in replication initiation and partitioning. The remaining 3 predicted proteins were hypothetical (Table 3.2). An additional protein encoded on p09022 had 94% amino acid identity to a hypothetical protein of plasmid p0471 from an uncharacterized marine bacterial host (3).

Table 3.2. CDSs of plasmid p09022

CDS	Coordinates	Predicted aa length	Protein homolog	% Identity (Similarity)	Phage/ Plasmid	Organism	Accession No.	e-value
1	186-1	62	No homolog					
2	1209-241	323	Hypothetical protein VFA0519 (299 aa)	52 (68)	*	*Vibrio fischeri*	YP_206477	1e-78
3	1479-1610	44	No homolog					
4	2335-1631	235	Hypothetical protein, COG: 3440 restriction endonuclease (262 aa)	39 (55)		*Mariprofundus ferrooxydans*	ZP_01451818	2e-17
5	2615-2484	44	No homolog					
6	2776-3501	242	Hypothetical protein VP1272 (241 aa)	80 (89)		*Vibrio parahaemolyticus*	NP_797651	2e-109
7	3674-4939	422	Site-specific methyltransferase cytosine-directed (461 aa)	62 (78)	*	*Salmonella sp.*	A32008	2e-143
8	5017-5766	250	Type II restriction endonuclease (230 aa)	42 (59)		*Escherichia coli*	CAA57630	6e-26
9	6971-5853	373	No homolog					
10	7036-7365	110	No homolog					
11	7356-7652	99	No homolog					
12	8156-7863	98	No homolog					
13	9904-8945	320	Type II restriction endonuclease (272 aa)	42 (59)		*Acidovorax sp.*	YP_985143	5e-20
14	10449-9907	181	Resolvase (179 aa)	73 (87)	*	*Psychromonas ingrahamii*	YP_941712	2e-68
15	12645-10645	666	No homolog					
16	13576-12635	314	No homolog					
17	14253-13573	227	No homolog					
18	14973-14368	202	Site-specific recombinase (198 aa)	85 (90)	*	*Vibrio cholerae*	AAX89421	4e-84
19	16656-15580	359	Replication initiation protein (230 aa)	94 (97)	pKA1	*Vibrio cholerae*	AAW51295	2e-114
20	17599-18321	241	Hypothetical protein SOA0007 (240 aa)	87 (94)	pKA1	*Vibrio cholerae*	AAW51296	7e-117
21	18596-19060	155	Hypothetical protein SOA0008 (154 aa)	91 (94)	pKA1	*Vibrio cholerae*	AAW51297	2e-68
22	19057-19575	173	Hypothetical protein SOA0010 (172 aa)	87 (92)	pKA1	*Vibrio cholerae*	AAW51298	2e-81
23	19791-21068	426	No homolog					
24	21920-21093	276	Hypothetical protein V12B01 (375 aa)	50 (67)		*Vibrio splendidus*	ZP_00990407	6e-30
25	22047-23009	321	Hypothetical protein (291 aa)	28 (48)		*Photobacterium profundum*	YP_129917	2e-21
26	24318-23068	417	Hypothetical protein (357 aa)	28 (53)		*Alvinella pompejana*	AAQ75130	1e-25
27	25457-24387	357	ParB (356 aa)	83 (94)	*	*Vibrio cholerae*	AAW51292	6e-151
28	26656-25457	400	ParA (399 aa)	88 (92)	*	*Vibrio cholerae*	AAW51293	0
29	27124-28626	501	No homolog					
30	28635-29528	298	Hypothetical protein (306 aa)	94 (96)	p0471	Gram negative	AAN74629	7e-158
31	29752-30726	325	Hypothetical outer membrane protein (1004 aa)	30 (45)		*Burkholderia ambifaria*	ZP_01559248	2e-04
32	31006-30815	64	No homolog					

*Denotes a CDS with similarity to a phage-encoded protein as determined by Prophage Finder (9).

Identification of P1-like proteins on *Vibrio* plasmid p0908. The phage P1 of Enterobacteria has been isolated from enteric bacteria (69) and has been shown to infect diverse bacteria under certain laboratory conditions (69). P1-like proteins and evidence of intact P1 phage have been identified in freshwater (8); however, to date none have been identified in marine systems. We identified 23 CDSs on p0908 with similarity to P1-encoded proteins and 20 additional proteins with similarity to other phage-encoded proteins. The remaining 16 predicted proteins were similar to chromosomally or plasmid-encoded proteins and 41 with no similarity to previously characterized proteins. The 23 CDSs of p0908 encoding P1-like proteins also occur in the same genomic arrangement as reported for the P1 genome, with a few differences possibly due to rearrangements (Figure 3.2) (60). The majority of these P1-like proteins (20 of 23) exhibited the same direction of transcription (Figure 3.2). The G+C contents of CDSs encoding the P1-like proteins (47-54%) were more similar to the overall G+C contents of p0908 (49.2%), P1 (49%)(60), and the *E. coli* host (50%)(12), than *Vibrio* chromosomes (39-47%) (22, 46, 61, 90).

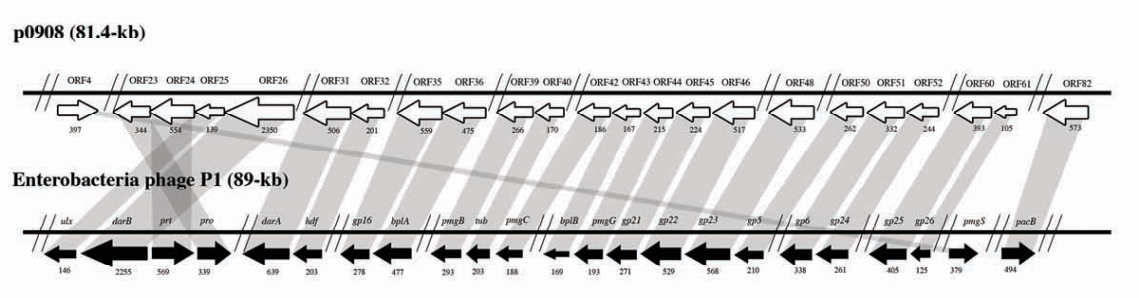


Figure 3.2. Genetic organization and amino acid conservation of predicted CDSs of p0908 compared to CDSs of Enterobacteria phage P1. Shading indicates regions with amino acid similarity, while the protein lengths (aa) are designated under each arrow and approximated by the arrow size. The orientation of each arrow indicates the direction of transcription. Vertical lines indicate the presence of additional genes that are not shown.

Of the proteins encoded on p0908 with similarity to P1 proteins there were 16 structural proteins, 6 antirestriction and head-processing proteins, and 1 involved in DNA packaging (Table 3.2). The structural proteins included those described as baseplate and tail tube (gp16, BplA, PmgB, Tub, PmgC, BplB, PmgG, gp5, gp6, gp24, gp25, gp26), sheath (gp21, gp22), and head (PmgS, gp23) components. The P1-like proteins involved in antirestriction and head-processing include DarA and DarB (60). Antirestriction proteins such as DarA and DarB prevent damage of phage DNA by host restriction enzymes (60). Identification of a protein encoded on p0908 with similarity to DarA strongly suggests p0908 may have acquired the P1-like genes from a P1 phage since DarA was shown to be unique to P1 (60). An additional indication of gene exchange between P1 and p0908 is the presence of an CDS encoding a protein similar to DarB. The predicted protein of CDS26 (2,350 aa) is comparable to DarB, which is the largest P1-encoded protein (2,255 aa) (60) (Table 3.2). Most of the P1-like genes of p0908

encode proteins for phage structure such as *tub* encoding a tail protein and *pro* involved in cleavage of head proteins during phage formation (60). P1 genes involved in prophage addiction, *phd* and *doc* (26), were noticeably absent from p0908. Although there were many phage structural proteins encoded on p0908 it is unlikely this is a functional phage as critical proteins for packaging and dispersal were absent. These included *lydA* and *lydB* encoding a holin and antiholin for host cell lysis (60). Of the proteins known to be required for functional packaging PacB was identified; however, PacA was absent. The gene encoding PacA includes the *pac* cleavage site, which is cleaved by the pacase enzyme composed of PacA and PacB proteins (60). A few proteins of KVP40, a T4-like phage, were similar to proteins of P1; however, this similarity was attributed to the relatedness of P1 and T4, both of the viral family *Myoviridae* (67). Two of these shared proteins were identified on p0908, BplA and Tub (67); however, numerous additional proteins similar to those encoded in the P1 genome were identified on p0908 that were not present on KVP40. Also encoded on p0908 are integrase-like proteins indicating the potential for integration of this element into a host chromosome (Table 3.3). Although vibriophage have been characterized with similarity to T4 (67), T7 (42), and P2 phage (77), none have been characterized with similarity to P1 (60). The P1 integrase (*cre*) was identified in bacterial lysogens from a freshwater pond indicating the presence of P1 in a freshwater environment (8). A viral metagenomic study produced two sequences with similarity to P1 PacA and PacB in estuarine waters of Southern California (15). To our knowledge no additional P1-like sequences or nearly complete P1 genomes have been identified from marine environments. The complete P1 sequence was finished after the viral metagenome was performed therefore some of the P1 genes may have not been

identified in the viral metagenome; however, a recent comparison of the P1 nucleotide sequence to the viral metagenome database primarily yielded hits to prophage from fish ponds (personal communication; Forest Rohwer). This indicates that additional P1 genes were not present in the marine viral metagenome. The pathogenic nature of some *Vibrio* spp. and possible residence in the gut may have facilitated a *Vibrio* MGE to exchange genes with P1 of an enteric bacterium resulting in an element such as p0908.

Table 3.3. CDSs of plasmid p0908

CDS	Coordinates	Predicted aa length	Protein homolog	% Identity (Similarity)	Phage/ Plasmids	Organism	Accession No.	e-value
1	512-3	170	Phage integrase	38 (54)		*Pseudomonas putida*	NP_746019	4e-20
2	600-908	102	No homolog					
3	984-1181	66	No homolog					
4	1237-2427	397	Hypothetical protein/PmgS (379 aa)	32 (56)	P1*	*Yersinia pseudotuberculosis*	YP_070308	8e-57
5	2477-2995	172	No homolog					
6	3087-3905	273	Hypothetical protein (190 aa)	34 (49)	*	*Escherichia coli*	ZP_00711985	7e-18
7	3962-4285	108	Hypothetical protein (114 aa)	32 (56)	*	*Burkholderia vietnamiensis*	ZP_00427437	6e-12
8	4343-4588	82	No homolog					
9	5166-4630	179	Hypothetical protein (122 aa)	42 (58)		*Vibrio alginolyticus*	ZP_01259570	4e-15
10	5718-5179	180	No homolog					
11	6745-6395	117	No homolog					
12	7597-6809	263	Hypothetical protein/trfA (285 aa)	35 (52)		*Burkholderia cenocepacia*	ZP_00462381	8e-28
13	8499-7588	304	Recombination associated protein (304 aa)	65 (79)	*	*Vibrio vulnificus*	NP_933515	1e-110
14	8946-8527	140	No homolog					
15	9415-9134	94	No homolog					
16	9621-9986	122	Hypothetical protein (100 aa)	43 (65)	CP-933N*	*Escherichia coli*	NP_287269	3e-13
17	9976-10338	121	No homolog					
18	10356-10616	87	No homolog					
19	10618-11460	281	Hypothetical DNA adenine methylase (263 aa)	36 (53)	*	*Synechococcus elongatus*	YP_400807	1e-37
20	12652-11615	346	Replication initiation protein (240 aa)	40 (59)	pKA1	*Vibrio cholerae*	AAW51295	2e-34
21	13354-14013	220	Partitioning protein, ParA (209 aa)	38 (61)	*	*Moraxella bovis*	BAD83724	5e-32
22	14006-14344	113	No homolog					
23	15495-14464	344	Pro (339 aa)	32 (54)	P1*	*Enterobacteria/ Escherichia coli*	YP_006481	1e-28
24	17149-15488	554	Prt (569 aa)	34 (57)	P1*	*Enterobacteria/ Escherichia coli*	YP_006480	2e-77
25	17704-17288	139	Ulx (146 aa)	32 (54)	P1*	*Enterobacteria/ Escherichia coli*	YP_006478	1e-09
26	24841-17792	2350	DarB (2255 aa)	34 (52)	P1*	*Enterobacteria/ Escherichia coli*	YP_006479	0
27	25179-25790	204	Hypothetical protein (229 aa)	30 (52)		*Photorhabdus luminescens*	NP_928281	4e-06
28	25956-26519	188	No homolog					
29	28334-26565	590	Hypothetical protein (1696 aa)	26 (48)	*	*Myxococcus xanthus*	YP_632657	5e-11
30	28620-28309	104	No homolog					
31	30143-28626	506	DarA (639 aa)	27 (49)	P1*	*Enterobacteria/ Escherichia coli*	YP_006494	2e-18
32	30742-30140	201	Hdf (203 aa)	26 (49)	P1*	*Enterobacteria/ Escherichia coli*	YP_006495	2e-07
33	31329-30853	159	No homolog		*			
34	31848-31411	146	No homolog		*			
35	33529-31853	559	gp16 (278 aa)	27 (46)	P1*	*Enterobacteria/ Escherichia coli*	YP_006504	3e-11
36	35065-33641	475	BplA (477 aa)	30 (49)	P1*	*Enterobacteria/ Escherichia coli*	YP_006505	3e-56
37	35413-35069	115	No homolog					
38	38613-35416	1066	Hypothetical protein (745 aa)	54 (75)		*Pseudomonas aeruginosa*	NP_249316	4e-08
39	39540-38743	266	PmgB (293 aa)	32 (49)	P1*	*Enterobacteria/ Escherichia coli*	YP_006508	2e-23
40	40052-39543	170	Tub (203 aa)	34 (53)	P1*	*Enterobacteria/ Escherichia coli*	YP_006509	3e-13
41	40401-40042	120	No homolog					
42	40969-40412	186	PmgC (188 aa)	37 (51)	P1*	*Enterobacteria/ Escherichia coli*	YP_006510	1e-19
43	41466-40966	167	BplB (169 aa)	34 (51)	P1*	*Enterobacteria/ Escherichia coli*	YP_006522	2e-18
44	42124-41480	215	PmgG (193 aa)	35 (49)	P1*	*Enterobacteria/ Escherichia coli*	YP_006523	3e-22
45	42795-42124	224	gp21 (271 aa)	32 (56)	P1*	*Enterobacteria/ Escherichia coli*	YP_006524	1e-16
46	44356-42806	517	gp22 (529 aa)	32 (49)	P1*	*Enterobacteria/ Escherichia coli*	YP_006525	8e-63
47	44866-44363	168	No homolog					
48	46534-44936	533	gp23 (568 aa)	31 (52)	P1*	*Enterobacteria/ Escherichia coli*	YP_006526	9e-73
49	46988-46788	67	Hypothetical protein V12B01 (65 aa)	46 (61)		*Vibrio splendidus*	ZP_00992067	0.32
50	47773-46988	262	gp5 (210 aa)	31 (47)	P1*	*Enterobacteria/ Escherichia coli*	YP_006545	3e-20
51	48791-47796	332	gp6 (338 aa)	26 (50)	P1*	*Enterobacteria/ Escherichia coli*	YP_006546	8e-29
52	49519-48788	244	gp24 (261 aa)	24 (42)	P1	*Enterobacteria/ Escherichia coli*	YP_006547	0.2
53	49888-49592	99	ORF38 (91 aa)	49 (70)	VHML*	*Vibrio harveyi*	NP_758930	6e-17
54	51158-49890	423	RNA-directed DNA polymerase (424 aa), ORF37	72 (83)	*	*Shewanella* sp., VHML	YP_963843	2e-168
55	51508-51155	118	Hypothetical protein (117 aa)	78 (90)		*Shewanella* sp.	YP_963844	7E-43
56	52805-51576	410	ORF35 (410 aa)	60 (73)	VHML*	*Vibrio harveyi*	NP_758928	7E-131
57	53023-52826	66	No homolog					
58	53949-53026	308	No homolog					
59	54686-53946	247	No homolog					
60	55864-54686	393	gp25 (405 aa)	34 (52)	P1*	*Enterobacteria/ Escherichia coli*	YP_006549	1e-32
61	56175-55861	105	gp26 (125 aa)	30 (55)	P1*	*Enterobacteria/ Escherichia coli*	YP_006550	6e-05
62	56798-56172	209	No homolog		*			
63	57113-57598	162	Hypothetical protein (176 aa)	28 (44)		*Parvibaculum lavamentivorans*	ZP_01659413	9e-04

Identification of additional phage-like proteins on *Vibrio* plasmids. The prevalence of P1-like proteins on p0908 led us to examine the occurrence of additional proteins typical of phage on all available *Vibrio* plasmids. Several non-P1 phage-like proteins were identified on the plasmids and in some cases had conserved gene order as well as amino acid identity. BLAST (5, 91) searches of the plasmid genomes to a phage-only sequence database using Prophage Finder (13) identified plasmid CDSs encoding proteins similar to phage proteins. There were 43, 5, and 6 predicted proteins with similarity to phage proteins encoded by CDSs of p0908, p23023, and p09022, respectively (Tables 3.2-3.4). Functions assigned to these proteins included replication, partitioning, transcriptional regulation, methylation, and recombination.

Of the 16 complete *Vibrio* plasmid sequences currently available (pES100, pYJ106, pJM1, pEIB1, pPS41, pSA19, pSIO1, pTC68, pVS43/pVS54, pES213, pTLC, pC4602-1, pC4602-2, pMP-1, and pR99) we detected the highest frequency of phage proteins on plasmids described in this study. These proteins encoded by CDSs of p0908, p23023, and p09022 represented 43, 8, and 20% of the total predicted CDSs, respectively. In contrast, the other large plasmid sequences available, pES100 (45.8-kb)(90), pJM1 (65-kb)(33), pEIB1 (66.1-kb)(115), and pC4602-1 (56.6-kb), pC4602-2 (66.9-kb), pR99 (68.4-kb), and pYJ106 (48.5-kb)(22), isolated from *V. fischeri, V. anguillarum,* and *V. vulnificus*; respectively, encoded proteins with similarity to phage proteins that comprised between 4 -14% of the predicted CDSs. Of the remaining nine plasmids, all less than 8-kb in size, pMP-1 (7.6-kb) had 3 proteins and pTLC (4.7-kb) (89), pVS43 (4.3-kb) and pVS54 (5.4-kb) isolated from a single strain of *V. salmonicida* encoded a protein with similarity to a protein associated with a phage.

The additional non-P1 phage proteins identified by BLAST analysis included recombinases, transcriptional regulators, transposases, and hypothetical proteins. Specifically, CDSs 53-56 of p0908 encoded proteins with 49-72% amino acid identity to CDSs 35-38 of phage VHML of *V. harveyi* (Table 3.2) (77). The comparable predicted amino acid size, identical gene order, and high amino acid identity of these proteins suggests recombination between *Vibrio* phage and plasmid elements. Additional proteins identified on phage and other plasmids include hypothetical proteins with a helix-turn-helix (H-T-H) motif. The H-T-H motif is typical of transcriptional regulators and other proteins with DNA-binding activity (9, 102). The amino acid sequence of CDS27 of p0908 is one example with 30% amino acid identity to a hypothetical protein of *Photorhabdus luminescens* and 25% identity to the *luxR* of *V. parahaemolyticus*. CDSs 10 and 21 of p23023 were shown to have similar H-T-H motifs. Hypothetical proteins with H-T-H motifs were also reported for predicted proteins of vibriophage VP16C and VP16T (93). To our knowledge the function and role of these putative transcriptional regulators for plasmid or phage stability has not been characterized.

Identification of conserved *Vibrio* chromosomal genes on *Vibrio* plasmids. The three plasmids examined in this study encoded numerous CDSs with significant amino acid identity (33-81%) to chromosomally encoded genes of *Vibrios* (Tables 3.2-3.4). To our knowledge, these chromosomally encoded genes have exclusively been identified on chromosomes and not on MGEs. Among those with significant amino acid identity were RecA (81%), a nucleoid-associated protein NdpA (65%), a Type IV helicase UvrD (65%), and a number of hypothetical proteins (50-80%). A RecA protein was previously reported on plasmid pNP40 (65-kb) from *Lactococcus lactis* (38); however, none have

been identified to date on *Vibrio* plasmids. Also, the plasmid-encoded RecA described in this study has greater protein identity (81%) to *Vibrio* RecAs than the lactococcal plasmid RecA had to other characterized lactococcal RecAs (45% aa identity) (38).

Table 3.4. CDSs of plasmid p23023

ORF	Coordinates	Predicted aa length	Protein homolog	% Identity (Similarity)	Phage/ Plasmid	Organism	Accession no.	e-value
1	307-2	101	No homolog					
2	536-315	74	No homolog					
3	1431-550	294	No homolog					
4	2377-1424	318	No homolog					
5	3174-2389	262	No homolog					
6	5756-3219	846	VirB4, Type IV secretion (790 aa)	26 (41)		*Xanthobacter autotrophicus*	ZP_01200724	4e-19
7	6083-5760	108	No homolog					
8	6456-6076	127	No homolog					
9	6705-6490	72	No homolog					
10	7564-8187	208	Hypothetical, H-T-H motif (203 aa)	39 (60)	*	*Shewanella baltica*	ZP_00584297	2e-30
11	8644-9489	282	Hypothetical protein, VVP68 (292 aa)	58 (75)		*Vibrio vulnificus*	NP_932268	9e-79
12	9691-9872	60	No homolog					
13	10484-10269	72	No homolog					
14	10829-10533	99	Hypothetical, Histone-like DNA-binding protein (91 aa)	31 (57)	*	*Oceanobacter* sp.	ZP_01306426	3e-06
15	11317-10928	130	No homolog					
16	11862-11314	183	Hypothetical protein (290 aa)	48 (58)		*Geobacter uraniumreducens*	ZP_01140337	4e-03
17	12484-12666	61	No homolog					
18	12734-12994	87	Hypothetical protein (86 aa)	56 (80)		Delta Proteobacterium	ZP_01291112	2e-26
19	12994-13260	89	Hypothetical protein (104 aa)	43 (56)		*Rhodospirillum rubrum*	YP_426088	8e-11
20	13279-13401	41	No homolog					
21	13572-13913	114	Hypothetical, H-T-H motif (118 aa)	75 (87)	*	*Marinomonas* sp.	ZP_01076798	1e-31
22	13942-14349	136	No homolog					
23	14497-14655	53	No homolog					
24	16059-14965	365	recA (349 aa)	81 (90)	*	*Vibrio angustum*	NP_798929	5e-155
25	16289-16687	133	No homolog					
26	16684-16944	87	No homolog					
27	17106-17411	102	No homolog					
28	17673-18086	138	Hypothetical protein (146 aa)	35 (52)		*Citrobacter freundii*	NP_775003	7e-13
29	18475-19359	295	Antirestriction protein (326 aa)	56 (70)		*Yersinia pseudotuberculosis*	YP_068542	4e-88
30	21829-19616	738	Hypothetical TraE (730 aa)	43 (57)		*Achromobacter denitrificans*	NP_990926	7e-154
31	22620-21883	246	Hypothetical Mobilization protein (248 aa)	37 (56)		*Erwinia carotovora*	YP_049708	9e-39
32	23174-22617	186	No homolog					
33	25205-23187	673	TriK, Mobilization protein (613 aa)	54 (72)		*Yersinia enterocolitica*	CAD58575	3e-173
34	26561-26154	136	DNA-binding protein, COG:2916 (135 aa)	33 (57)		*Actinobacillus pleuropneumoniae*	ZP_00134899	8e-14
35	26598-26804	69	No homolog					
36	27767-26856	304	Hypothetical protein (295 aa)	39 (54)		*Shewanella baltica*	ZP_01437045	2e-51
37	28068-27739	110	No homolog					
38	29090-28710	127	No homolog					
39	29631-29239	131	No homolog					
40	30322-29720	201	Hypothetical protein (177 aa)	37 (55)		*Proteus vulgaris*	NP_640181	2e-07
41	30634-30527	36	No homolog					
42	31261-30743	173	No homolog					
43	31589-31245	115	Hypothetical protein (113 aa)	31 (49)		*Nitrosomonas eutropha*	YP_746361	4e-05
44	32617-31631	329	Hypothetical protein (432 aa)	39 (60)		*Vibrio cholerae*	ZP_01483941	1e-51
45	33500-32661	280	Hypothetical protein (155 aa)	26 (42)		*Vibrio fischeri*	YP_207178	1e-06
46	34003-33497	169	Lytic transglycosylase (329 aa)	51 (70)	*	*Shewanella frigidimarina*	YP_752524	5e-25
47	34749-33982	256	Hypothetical protein VFB37 (210 aa)	29 (47)		*Vibrio fischeri*	YP_207167	3e-07
48	37528-34943	862	UvrD-like helicase IV (962 aa)	65 (79)		*Vibrio splendidus*	ZP_00990408	0
49	39347-38079	423	Hypothetical protein (352 aa)	27 (41)		*Photobacterium profundum*	YP_133279	2e-23
50	40046-39534	171	No homolog					
51	41034-40084	317	Hypothetical protein (318 aa)	37 (60)		*Vibrio fischeri*	YP_207163	8e-48
52	41556-41053	168	Hypothetical transporter (159 aa)	18 (44)		*Vibrio fischeri*	YP_207162	3e-03
53	42045-41557	163	Hypothetical protein VFB31 (164 aa)	44 (62)		*Vibrio fischeri*	YP_207161	2e-31
54	43095-42061	345	Hypothetical protein VFB30 (338 aa)	33 (57)		*Vibrio fischeri*	YP_207160	5e-52
55	44738-43104	545	Hypothetical Type II secretion protein (598 aa)	48 (66)		*Shewanella baltica*	ZP_01437079	4e-136
56	45487-44738	250	Hypothetical protein (374 aa)	25 (45)		*Burkholderia vietnamiensis*	ZP_00423077	3e-04
57	46610-45471	380	Hypothetical protein (370 aa)	22 (41)		*Shewanella baltica*	ZP_01437077	2e-10
58	48134-46623	504	Hypothetical protein (488 aa)	46 (67)		*Photobacterium profundum*	YP_133269	8e-126
59	48880-48134	249	Hypothetical protein VFB25 (243 aa)	36 (56)		*Vibrio fischeri*	YP_207155	1e-37
60	49238-48873	122	Hypothetical protein VFB24 (117 aa)	49 (66)		*Vibrio fischeri*	YP_207154	1e-22
61	50529-49312	406	Hypothetical conjugal transfer protein (415 aa)	29 (50)		*Desulfotalea psychrophila*	YP_066874	1e-15
62	51371-50529	281	No homolog					
63	52124-51396	243	Hypothetical conjugal transfer protein (227 aa)	23 (39)		*Erwinia carotovora*	YP_049716	2e-03
64	52353-52132	74	No homolog					
65	52525-52367	29	No homolog					

The RecA (CDS24) encoded on p23023 (Table 3.4) was more similar to RecA of other *Vibrios* (81% aa identity) than to that of related gamma-proteobacteria such as *Photobacterium* spp. (Figure 3.3). This indicates the plasmid-encoded *recA* was likely from a *Vibrio* host. Sequence alignment of the predicted amino acid sequence of CDS24 to RecA sequences of *V. mediterranei*, *V. splendidus, V. parahaemolyticus,* and *E. coli* shows the extent of conservation of CDS24 to *Vibrio* RecAs (Figure 3.4). The RecA signature motif characteristic of RecA proteins is present in all the aligned sequences (Figure 3.4)(9). Also, the P-loop motif for ATP binding, which is characteristic of ATPase-like proteins is present in CDS24 (7, 113) (Figure 3.4). The DNA-binding loops, L1 and L2, involved in double-stranded and single-stranded DNA-binding (38, 98); respectively, are also present in CDS24. The DNA-binding loop L1 of CDS24 is identical to the same motif found in other RecAs (38, 98). In contrast, loop L2 contains a gap and two other amino acid changes that may alter the single-stranded DNA-binding activity of the protein encoded by CDS24. Based on sequence analysis of CDS24 the predicted protein likely has the recombinase (24) and proteolytic cleavage activity that has been characterized to date for other RecAs (35, 55).

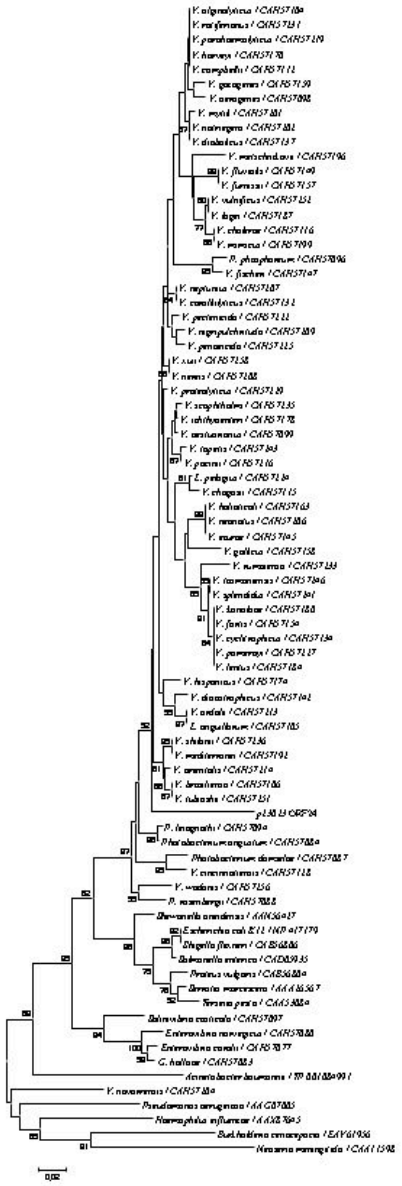


Figure 3.3. Phylogenetic comparison of the predicted amino acid sequence encoded by CDS24 of p23023 with RecA amino acid sequences representing most *Vibrio* spp. and additional members of the *Vibrionaceae* available in Genbank as of March 2007. Distantly related proteobacteria were included as outgroups. A neighbor-joining tree was constructed in MEGA (56) using the Poisson correction (73). Bootstrap values were generated over 1,000 replications and are indicated where ≥ 50.

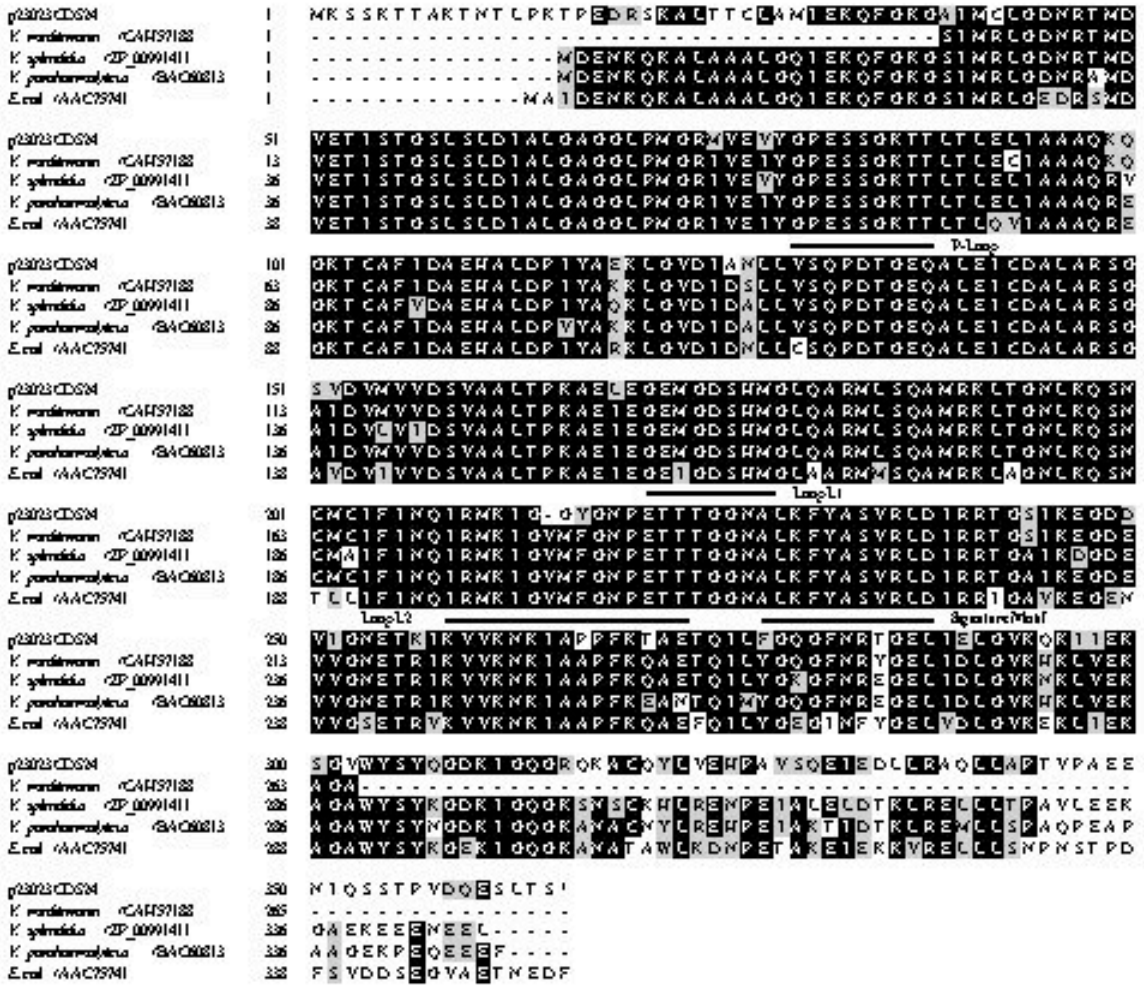


Figure 3.4. Sequence alignment of RecA amino acid sequences of CDS24 from p23023 with those sequences encoded on the genomes of *V. splendidus, V. parahaemolyticus,* and *E. coli*. The *V. mediterranei* RecA sequence was a partial sequence obtained from a multilocus sequence analysis of *Vibrio* spp. (106). The RecA signature motif (7), P-loop motif (Walker A motif) involved in ATP binding (9), loop L1 motif involved in binding of double-stranded DNA (38, 98), and the loop L2 motif for binding single-stranded DNA (38, 98) are all indicated.

RecA protein sequences have been shown for some bacterial species to provide greater resolution than phylogenetic analysis of an equal number of 16S rRNA sequences (35). Phylogenetic assessment of *Vibrios* has previously demonstrated *recA* nucleotide sequences can be used as an alternate phylogenetic marker to 16S rRNA (97, 105, 106). Several studies revealed considerable sequence variation (0-6%) of *recA* for certain *Vibrio* spp. (106) with as low as 94% *recA* nucleotide identity within a species. In contrast, *Photobacterium* spp., also within the *Vibrionaceae,* had less than 94% *recA* identity to the closest related *Vibrio recA*. Overall, CDS24 is highly conserved compared to other RecA sequences; however, the N- and C- terminal regions are significantly diverged (Figure 3.4). Sequence analysis of RecAs from diverse bacteria have revealed the majority of the protein to be highly conserved while the N- and C-termini were significantly variable (35). The observed sequence divergence of the termini of the predicted protein sequence of CDS24 may have occurred by recombination with alleles having greater sequence divergence after the sequence was acquired by the plasmid. Alternately, selection pressure for a specialized role of the plasmid-encoded RecA for plasmid stability or uncharacterized functions may have lead to the sequence divergence in the terminal regions.

This study is the first report of a *recA* encoded on a plasmid isolated from a *Vibrio* host. The potential for horizontal transmission of *recA* by *Vibrio* plasmids raises questions of whether *recA* provides reliable resolution for discriminating between related *Vibrio* spp. (105) or determining the extent of O-antigen gene exchange (97). Additional plasmid-encoded proteins with similarity to conserved chromosomal genes include the nucleoid associated protein NdpA and a UvrD-like helicase. The plasmid-encoded NdpA

reported here, CDS97 of p0908, had 65% amino acid identity (79% similarity) to NdpA of *V. vulnificus* (Table 3.3). Also, CDS48 of p23023 had 65% amino acid identity (79% similarity) to a UvrD-like helicase of *Vibrio splendidus* (Table 3.4). Recombination of host and plasmid-encoded *uvrD*-like genes may increase diversity of *uvrD*, potentially disrupting function of the host protein. To our knowledge this is the first description of an NdpA-like protein and a UvrD-like helicase encoded on a plasmid. These results indicate *Vibrio* plasmids may be involved in horizontal dissemination of conserved genes, such as *recA* and *uvrD* both involved in host adaptive responses. Further investigation of the diversity encoded by *Vibrio* plasmids would be necessary to determine the extent that these elements transfer conserved genomic regions among diverse *Vibrio* spp.

In addition to high amino acid identity, several proteins had an identical gene order in the plasmid as found in the *Vibrio* chromosomes. Specifically, CDSs 51-54 and 59-60 of p23023 encoded predicted proteins with amino acid identity and conserved gene order to that reported for hypothetical proteins of *V. fischeri* (Table 3.4).

This study provides evidence for a role of *Vibrio* plasmids in gene exchange among diverse *Vibrio* spp. as evident by the gene content and unique genomic signatures of *Vibrio* plasmids relative to *Vibrio* chromosomes. Identification of P1-like proteins and other phage-like proteins on *Vibrio* plasmids support the mosaicism of *Vibrio* MGEs and the potential for recombination between *Vibrio* plasmids and phage. The considerable diversity of *recA* among strains of certain *Vibrio* spp. may be facilitated by recombination of plasmid-encoded genes such as the p23023 *recA*. Further studies into the genetic diversity of *Vibrio* plasmids as well as their potential host range are needed to better understand the evolution of MGEs and their role in diversification of *Vibrio* spp. This

will serve as the basis for future molecular investigations into the role of plasmids for unique phenotypes promoting adaptation to fluctuating environmental conditions and the potential emergence of pathogens.

Acknowledgements

I would like to thank J. Eisen and D. Wu for their collaboration on the plasmid sequencing and characterization. I would like to thank M. Barnstead, S. van Aken, G. Pai, and M. B. Craven for coordinating the library construction and sequencing of the plasmids. Also, I would like to thank K. Ayanbule for technical suggestions on the annotation.

Part 2. Diversity and Distribution of *Vibrio parahaemolyticus* Plasmids and Their Role in the Evolution of Clinical and Environmental Strains

Abstract

Vibrios have diverse mobile genetic elements (MGEs) including phage, transposable elements, and plasmids that act as agents of horizontal gene transfer (HGT). To date, little is known of the distribution and diversity of *V. parahaemolyticus* MGEs and their role in the horizontal transfer of genes conferring pathogenicity or environmental survival and niche expansion. In the current study, we characterized the sequences of a plasmid and a prophage isolated from a *V. parahaemolyticus* environmental strain. Sequence characterization of the 28.8-kb plasmid, p22702B, from a *V. parahaemolyticus* environmental strain revealed the role of this plasmid in horizontal transfer of genes associated with *V. parahaemolyticus* clonal pandemic strains. *V. parahaemolyticus* plasmid p22702B encoded a bacteriocin with 98% amino acid identity to *vpa1263* of the genomic island VPaI-6 that is characteristic of the pandemic strains. In addition, we show that p22702B belongs to a plasmid family with sizes ranging from 30-kb to 100-kb that are present among *V. parahaemolyticus, V. harveyi,* and *V. campbellii* strains isolated from environmental samples. Furthermore, we identified an 86-kb plasmid that is present among a *V. parahaemolyticus* O4:K12 clonal strain isolated from a disease-associated sample and two strains isolated from geographically separate environmental samples. This study shows the horizontal transfer genomic island genes of *V. parahaemolyticus* pandemic strains demonstrating a role of plasmids in the emergence of *V. parahaemolyticus* disease-causing strains from environmental populations.

Introduction

Vibrio parahaemolyticus is a halophilic marine bacterium and an opportunistic human pathogen that is responsible for the majority of the seafood-borne illnesses in the United States annually (1, 107). *V. parahaemolyticus* typically causes gastroenteritis and less commonly has been associated with wound infections upon exposure to water (36, 71, 107). Increasing reports of *V. parahaemolyticus*-related illnesses have been linked to the emergence of a clonal pandemic strain in Asia in 1995 and its subsequent spread worldwide (2, 30, 62, 109). The pandemic strains have been identified with 24 characteristic DNA regions that were likely acquired by horizontal gene transfer (HGT) (14). The *V. parahaemolyticus* O3:K6 pandemic strain RIMD2210633 genome possesses seven genomic islands designated VPaI-1 to VPaI-7 (49). The virulence mechanism of *V. parahaemolyticus* has not been characterized; however, strains with a diverse repertoire of potential virulence mechanisms have been isolated from illness-associated samples. Among the virulence-associated genes are the thermostable direct hemolysins (*tdh* and *trh*) and two type III secretion systems, T3SS1 and T3SS2, located on chromosomes I and II, respectively (61, 81). T3SS1 has been detected among all *V. parahaemolyticus* strains (44, 61, 81) while T3SS2 is present in the genomic island VPaI-7 with two copies of *tdh* that is present only in pandemic strains (44, 49, 81). Recently, a second T3SS2, designated T3SS2β, was identified as part of a genomic island that contains *trh* and is present in tdh^{-}/trh^{+} strains (78).

Molecular analysis of the diversity of *V. parahaemolyticus* strains showed that strains isolated from clinical and environmental sources have significant genetic diversity (62) compared to the clonal pandemic strains that are frequently associated with disease

outbreaks (14, 23). Of the extrachromosomal genetic elements that have been characterized from *V. parahaemolyticus,* the most is known about the filamentous vibriophage. Vibriophage are abundant among environmental strains (25-27); however, the contribution of phage and other genetic elements to the diversification of *V. parahaemolyticus* strains is not well understood. In particular, the diversity of plasmids associated with *V. parahaemolyticus* and their role in horizontal gene transfer has not been characterized. A 35-kb plasmid isolated from *V. parahaemolyticus* was shown to encode the virulence-associated thermostable direct hemolysin (*tdh*) (74). Previous studies examining the diversity of *V. parahaemolyticus* MGEs have focused on small plasmids (<10-kb) (GenBank) and vibriophage (20, 21, 72). A previous study that examined the frequency of *V. vulnificus* and *V. parahaemolyticus* strains that produced bacteriocin-like substances demonstrated that approximately 30% of *V. parahaemolyticus* environmental strains possess MGEs (94); however, to our knowledge no studies have examined the genetic diversity of *V. parahaemolyticus* plasmids.

Genomic islands have been identified in *V. cholerae, V. vulnificus,* and *V. parahaemolyticus* genomes and are often associated with innovative phenotypes such as pathogenicity or survival in adverse environments (88). Although genomic islands have been identified in numerous *Vibrio* spp., the frequency of gene exchange between genomic islands and other mobile genetic elements (MGEs) such as plasmids and phage is not known. Excision of the genomic islands to generate circular intermediates was not demonstrated for *V. parahaemolyticus* (49); however, three pathogenicity islands of *V. cholerae* were shown to excise producing circular intermediates (70). Another mechanism that may facilitate the transfer of the plasmids examined in this study is

natural transformation. *V. cholerae* has previously been shown in the presence of chitin and during nutrient limitation to naturally transform free DNA up to 44.9-kb (66, 68).

Vibrio spp. are known to survive extended starvation and other environmental stressors including changes in temperature and salinity (75, 76, 80, 99). Following periods of nutrient stress and low temperature, *Vibrio* spp. enter into a viable but non-culturable state (VBNC) from which they can be resuscitated (10, 79) as a mechanism to tolerate changes in temperature and availability of nutrients. *Vibrios* have been shown to coordinate their survival responses to nutrient limitation and predation by chemical communication (64); however, mechanisms of competition such as the production of killing factors have not been examined for *V. parahaemolyticus*. Bacteriocins have been shown to confer an advantage during survival and competition for limiting nutrients (48). The distribution of bacteriocins, association of bacteriocins with MGEs, and contribution of bacteriocins to survival has been examined for only a few *Vibrio* spp. *V. harveyi* (85), *V. vulnificus* (94), and *V. mediterranei* (18) have been shown to produce a substance that inhibits growth of *Vibrios* and additional marine bacteria. *V. parahaemolyticus* is inhibited by bacteriocins produced by other *Vibrios* (18, 94); however, to date *V. parahaemolyticus* strains have not been shown to produce an inhibitory substance. *V. harveyi* was shown to produce a bacteriocin-like inhibitory substance attributed to the presence of plasmid DNA (63). The plasmid-bearing *V. harveyi* exhibited better survival over an isogenic plasmid-free strain when grown in competition during nutrient limitation (48).

In this study, we characterized the distribution and diversity of plasmids associated with *V. parahaemolyticus* clinical and environmental strains. Sequence

analysis of the 28.8-kb plasmid, p22702B, that was isolated from a *V. parahaemolyticus* environmental strain showed that plasmids facilitate gene transfer between *V. parahaemolyticus* pandemic strains and environmental strains. This study provides the first characterization of the diversity of *V. parahaemolyticus* plasmids among clinical and environmental strains as well as the sequence characterization of a *V. parahaemolyticus* plasmid greater than 9-kb. The putative bacteriocin gene, *vpa1263,* present in the VPaI-6 genomic island and identified on p22702B with 98% amino acid identity was only detected among *V. parahaemolyticus* pandemic strains examined in this study.

Materials and Methods

Bacterial strains and media. *V. parahaemolyticus* strains of clinical origin were provided by the Centers for Disease Control and Prevention (CDC; Atlanta, GA). *V. parahaemolyticus* environmental strains were isolated from the saltmarsh rhizosphere of North Inlet, North Carolina (6) as well as saltmarsh sediment and water from Skidaway Island, GA and Appalachicola, FL (43). *V. parahaemolyticus* 22702 was isolated from coastal marine sediment samples from Sapelo Island, GA (28, 29) and identified by biochemical and phylogenetic analyses (29). Serotyping of *V. parahaemolyticus* 22702 was performed at the CDC as previously described (32, 36). Phylogenetic identification of *V. parahaemolyticus* 22702 was performed in a previous study (45). Additional environmental strains examined in this study were cultured from sediment and water samples from the saltmarsh of Skidaway Island, GA during September 2006 and 2007 (43). All marine bacteria were grown at 30˚C in M10 broth (43) supplemented with 1.8%

agar for plates while *E. coli* was grown in LB broth at 37˚C with shaking or on LB plates with 1.5% agar.

Plasmid isolation, sequencing, and analysis. Sequencing of the plasmids p22702B and p22702A was performed as previously described (45). Also, annotation and sequence analysis was performed as described (45). A modified Kieser acid phenol extraction method (54) was used to isolate extrachromosomal elements from chromosomal DNA. Briefly, *Vibrio* strains were grown overnight in 5 ml of LB or M10 broth at 30˚C with shaking. Cultures were transferred to falcon tubes and centrifuged at 7,000 rpm for 7 min at 4˚C and the supernatant discarded. Cell pellets were resuspended in 500 µl of lysis solution (25 mM Tris-HCl (pH 8), 25 mM EDTA (pH 8), 0.3 M sucrose, 0.02% bromocresol green, 2 mg/ml lysozyme) and transferred to 1.5 ml microcentrifuge tubes. Following incubation at 37˚C for 30 min 250 µl of a neutralization solution (0.3M NaOH, 2% SDS) was added and tubes were incubated at 60˚C for 30 min. Tubes were cooled to room temperature then 200 µl of an acid phenol solution (5 ml melted acid phenol, 5 ml unbuffered chloroform, 1 ml water, 5 mg 8-hydroxyquinoline) was added to each tube. The tubes were vigorously shaken until the solution because opaque then they were centrifuged at 14,000 rpm for 5 min at 4˚C. The upper translucent blue aqueous phase (approximately 70-80 µl) was transferred to a sterile microcentrifuge tube and stored at 4˚C. Extrachromosomal bands were visualized by running 80 µl of the extract on a 0.7% SeaKem LE agarose gel for 4-5 hours at 100V then stained and de-stained with ethidium bromide. Primers were designed from the p22702B and p09022 *rep* genes to screen for the presence of related *rep* genes in other strains. The predicted *rep* gene was sequenced

for all strains that tested positive for the p22702B/p09022 *rep* using primers listed in Table 3.5.

Table 3.5. Primers used in this study

Primer	Sequence (5'- 3')	Tm	Expected amplicon (bp)	Source
Plasmid Rep				
rep M13F	TGTAAAACGACGGCCAGT GCACGAGAAGCCGATATGTTTGCA	69	707	This study
rep M13R	CAGGAAACAGCTATGACC TCCAGCGGAACAGGAACTTGATAC	67	"	"
Phage Rep				
rstA M13F	TGTAAAACGACGGCCAGTCGATTGCACAAGCATGAGGA	68	906	"
rstA M13R	CAGGAAACAGCTATGACCGGATTTTCCCGTACTGACGA	66	"	"
Gene deletion				
*vpa1263*D1ApaI	GACTGGGCCCCCAACACCGACCAACTTCTT	68	900	"
*vpa1263*D2	TTACCCGCATCAGCTAACCACGTATCACTAACGAGTGCGA	68	"	"
*vpa1263*D3	TCGCACTCGTTAGTGATACGTGGTTAGCTGATGCGGGTAA	68	900	"
*vpa1263*D4SacI	GACTGAGCTCCTCGAGGAGCTTTGCCATCC	66	"	"
vpa1263 62F	CCAACACCGACCAACTTCTT	55	1,000	"
vpa1263 797R	AAGGGTGGAAGATCAGAGCA	56	"	"
RT-PCR				
rpoB 1110F	GTAGAAATCTACCGCATGATG	51	400	(101)
rpoB CM32b	CGGAACGGCCTGACGTTGCAT	63	"	"
vpa1263 expF	ACCACGGCGCAAAGCCGCGTTCGTGTC	70	312	This study
vpa1263 expR	CTT CTG GCG CAG GCG TAA CGG TGA CGT T	67	"	

Primers used to PCR screen for the presence of *vpa1260* and *vpa1263* are listed in Table 3.5. The PCR screen was performed with NEB Taq polymerase and standard reaction and cycle conditions. PCR amplification for sequencing was performed using NEB Phusion high-fidelity polymerase reaction mix with GC buffer and standard reaction and cycle conditions.

Sequence and phylogenetic analysis. Phylogenetic analysis was performed for a concatenation of the housekeeping genes *recA* and *gyrB* of select plasmid-bearing strains using primers developed in a previous study (39). A phylogeny was also constructed for the nucleotide sequences of the replication protein-encoding genes *repA* and *rstA* from plasmids and phage, respectively. Neighbor-joining trees were constructed in MEGA

using the Kimura 2-parameter model (56). Bootstrap values represent 1,000 replications and values ≥ 50 are shown.

Southern hybridization. Plasmids and other extrachromosomal elements were separated from chromosomal DNA using a modified Kieser extraction protocol (54). Localization of the replication protein-encoding gene, *repA,* to plasmid DNA was confirmed by southern hybridization. Probes were generated by PCR incorporation of dig-labeled dUTP (Roche; Indianapolis, IN) into amplicons with *V. parahaemolyticus* 22702 as template DNA and primers listed in Table 3.5. The hybridization was performed according to the manufacturer's suggested protocol. Briefly, the membrane was generated by running Kieser extractions on a 0.7% SeaKem LE agarose gel for 4 hours at 100 V. The gel was depurinated in 0.25 M HCl and transferred overnight to a positively charge membrane (Roche) using capillary transfer with 20X SSC buffer. The DNA was UV cross-linked to the membrane and the membrane was rinsed with sterile water then allowed to dry before use. 50 µl of labeled probe DNA was added to 7 ml of DIG-easy-hyb granules to generate the hybridization solution. The membrane was pre-hybridized for 2 hr with DIG-easy-hyb granules at 42˚C. The membrane was hybridized with the *rep* probe at 42˚C overnight in a Hybaid oven. The membrane with washed with low stringency buffer at room temperature and high stringency buffer at 65˚C for 15 min. The membrane was washed with blocking solution for three hours then incubated with anti-DIG antibody for 30 min. The membrane was detected with chemiluminescent substrate (Roche) and exposed to chemiluminescent film for 15 min.

Plasmid restriction endonuclease profiles. Plasmids were isolated as described above and plasmid DNA was separated from chromosomal DNA on a 1.0% SeaKem gold agarose

gel. Plasmid bands were sliced from the gel and an in-slice restriction endonuclease digestion was performed on each plasmid band by modification of a previous protocol (L. Williams and A. Summers; unpublished). Briefly, plasmid slices were placed in a microcentrifuge tube and 1 ml of 10:1 TE (10 mM Tris, 1 mM EDTA) was added. Tubes were incubated at 37˚C with shaking (200 rpm) for 30 min. The TE was completely removed from each tube and a total of 100 µl of restriction buffer (1X restriction buffer, 1X BSA, and water) was added to each tube. Buffer should completely cover each slice. Tubes were then incubated at 37˚C with shaking for 30 min. The restriction buffer was removed and 100 µl of new restriction buffer and 20-40 U of each enzyme was added. Plasmids were double-digested with PacI and SbfI at 37˚C for 12 hours. The restricted plasmid bands were run on a 1% SeaKem gold agarose gel using pulsed-field gel electrophoresis (PFGE) as previously described (82). Restriction buffer was removed and 200 µl of 0.5X TBE was added to each tube. The tubes were incubated at room temperature for 5 min. Restricted gel slices were affixed to the teeth of a 10-well 100 ml gel comb. Once slices had dried a 1% SeaKem gold agarose gel was poured incorporating the plasmid slices into the gel. The gel was run in 0.5X TBE on a BioRad Chefmapper with a low molecular weight of 1-kb and a high molecular weight of 200-kb for 15 hours at 6 V/cm. A low range PFG marker (NEB; Ipswich, MA) was included in the first and last lanes for band sizing. The gel was visualized by staining with ethidium bromide for 15 min and de-staining twice with water for 30 min.

Construction of Δ*vpa1263*. We deleted approximately 80% of the predicted coding region of *vpa1263* in RIMD2210633 using methods previously described (43) with the following modifications. Primers used for the gene deletion are listed in Table 3.5. For

previous gene deletions (43), we found that the frequency of spontaneous *sacB* mutants was high and instead we used pKAS46 (*rpsL*:kan:amp) as a suicide vector for deletion of *vpa1263* (95). This vector encodes *rpsL*, which confers sensitivity to streptomycin sulfate (strep) as a method counter-selection. Briefly, we isolated a strep resistant ($strep^r$) mutant of *V. parahaemolyticus* RIMD2210633 by plating 10^9 cells on LB strep (5 mg/ml). Primers listed in Table 3.5 were used to PCR amplify the ends of *vpa1263* with additional upstream and downstream DNA and to fuse these regions together into a single construct. The fusion amplicon was double-digested with the restriction endonucleases ApaI and SacI and ligated into ApaI- and SacI-digested pKAS46 as previously described (43). The ligation was transformed into *E. coli* β2155-λ pir and plated on LB containing ampicillin (amp; 100 μg/ml) and diaminopimelic acid (dap; 100 μg/ml). Transformants containing the assembled suicide vector were identified by PCR screening for the insert using primers D1ApaI and D2SacI. The assembled vector was further confirmed by mini-prep recovery of vector DNA and restriction endonuclease digestion with ApaI and SacI. β2155 containing the assembled suicide vector was grown overnight in LB supplemented with dap_{100} and amp_{100} and transferred to *V. parahaemolyticus* RIMD2210633 $strep^r$ by biparental-mating as previously described (43). *V. parahamolyticus* colonies that received the vector were isolated by plating on LB supplemented with kanamycin (100 μg/ml) and without dap to exclude growth of β2155, which is a dap auxotroph. Colonies that grew on the LB kan were patched to LB supplemented with $strep_{5000}$ and grown overnight. Colonies that grew on LB with strep were patched again to LB strep and grown overnight. The $strep^r$ colonies were then patched to LB with kan_{100} to confirm loss of the suicide vector. Colonies that were $strep^r$

and kan^{s} were PCR screened using primers 62F and 797R (Table 3.5) to identify an isolate that had the truncated 1-kb deletion amplicon resulting from deletion of 2.1-kb of the 2.6-kb coding region of *vpa1263*. Sequence analysis of the deletion amplicon was performed to confirm the deletion was in-frame. The Δ*vpa1263* colony was mini-prepped as described above to confirm that the suicide vector was gone.

Screen for bacteriocin-like inhibition. Bacteriocin production was investigated using a spot-on-lawn assay and a flip-agar assay that were previously used to investigate the bacteriocin production by *V. harveyi* (63, 85, 94). The bacteriocin screens were performed using tryptic soy broth or LB, which were supplemented with 1.5% agar for plates. Both Difco granulated agar and Difco Bacto agar were used to make plates. Briefly, the spot-on-lawn assay was performed by growing strains were grown overnight in tryptic soy broth (TSB) at 30˚C with shaking (200 rpm). Then 100 µl of an indicator was plated on tryptic soy agar (TSA) and incubated for 10 min at 30˚C. Following incubation, 10 µl of a producer was spotted on the plate and incubated for 24 hours at 30˚C.

The flip-agar assay was performed as described with modifications as follows (85). The producer strains were grown overnight in TSB. A single 10µl spot of each producer was put in the center of a TSA plate. The plates were then incubated for 3 days at 30˚C. A heat-sterilized spatula was used to loosen the edge of the agar and to flip it over into the lid of the plate. A single indicator strain was swabbed onto the exposed surface of the agar. Plates were incubated for 12-24 hours at 30˚C and the presence of inhibition zones was recorded.

RT-PCR analysis. Reverse transcription PCR analysis (RT-PCR) of the expression of *vpa1263* was performed as previously described (66). Briefly, 1 ml of strains that were grown overnight in LB at 30˚C with shaking was extracted using Purezol (BioRad; Hercules, CA). Following extraction according to the product guidelines, the total RNA was quantified and was treated with DNaseI (Ambion; Foster City, CA) according to the product guidelines. The DNase-digested RNA was purified using an RNeasy column (Qiagen; Valencia, CA) and was reverse transcribed using iscript (BioRad; Hercules, CA) with a total of 500 ng of RNA for each sample. PCR amplification to detect *vpa1263* was performed with 2 µl of cDNA for each sample. The presence of *rpoB* transcripts was analyzed as a control to confirm whether successful RNA extraction. The *vpa1263* and *rpoB* primers were used in a previous study (101) and are listed in Table 3.5.

Competition assay. The liquid competition assay was performed as previously described for *V. harveyi* (48). Strains were grown overnight in LB at 30˚C with shaking and inoculated 1:1 into fresh LB. The cultures were diluted and plated on LB and LB supplemented with streptomycin sulfate (Str) at the initial inoculation and again at 10 hours and 24 hours. The CFU/ml of the Δ*vpa1263* strain was determined as the number of colonies growing on the LB Str plates, while the CFU/ml of the wild-type strain *V. parahaemolyticus* RIMD2210633 was calculated as the difference of the total CFU/ml on LB plates and the CFU/ml on LB Str plates.

Accession numbers. The nucleotide sequences of p22702B and p22702A are pending GenBank submission. The *recA, gyrB, rep,* and *rstA* sequences are deposited in GenBank under accession numbers FJ596927-FJ596987. The *recA* sequence of *V.*

parahaemolyticus 22702 was previously deposited in GenBank under the accession number EU018456.

Results

Sequence characterization of the *V. parahaemolyticus* plamids p22702A and p22702B.

The environmental strain *V. parahaemolyticus* 22702 isolated from a GA saltmarsh had two extrachromosomal elements, p22702A (9.8-kb) and p22702B (28.8-kb). Nucleotide sequence analysis of p22702A (9.8-kb) revealed this element had an overall G+C content of 46% and encoded 15 protein-coding regions (CDSs). Among the CDSs identified for p22702A were 10 that encoded proteins with 91-100% amino acid identity to those of f237 (Figure 3.5, Table 3.6). An additional CDS encoded a hypothetical protein that had 77% amino acid identity to ORF2 of phage f237. Four of the CDSs with identity to f237 ORFs were duplicated within the p22702A sequence (Figure 3.5). Also, p22702A encoded three proteins with high amino acid identity to hypothetical proteins VP1550 (97%) and VP1559 (97%) of *V. parahaemolyticus* and a hypothetical protein of *Vibrio* sp. Ex25 (94%) (Table 3.6; Figure 3.5). Interestingly, proteins VP1551-VP1554 and VP1556-VP1558 are similar to the proteins of ORF1-ORF4 and ORF5-ORF7 of the vibriophage f237. In contrast, *vp1555* encoded a hypothetical protein not present in the f237 sequence (Table 3.6). The duplication of four CDSs of p22702A may have resulted from improper excision of the phage from the chromosome. The possibility of gene duplication in p22702A following improper excision from the chromosome is supported by the presence of the chromosomal flanking genes VP1550 and VP1559 indicating p22702A was likely integrated into the chromosome. The *V. parahaemolyticus*

RIMD2210633 chromosome I has a filamentous phage integrated between *vp1550* and *vp1559*. Numerous methods of screening for plasmid content of *V. parahaemolyticus* 22702 revealed the presence of a small band of approximately 10-kb. The frequency of this band in repeated trials was low indicating this element was likely integrated into the chromosome and occasionally excised to generate the replicative prophage identified in some replicates (data not shown).

In addition to the putative prophage p22702A, *V. parahaemolyticus* 22702 contained a 28.8-kb plasmid designated p22702B. Nucleotide sequence analysis revealed p22702B had 26 CDSs. The overall G+C content of p22702B is 42%, which is lower than the average G+C content (45%) of the genome of *V. parahaemolyticus* RIMD2210633 (61). Addiction systems identified on p22702B include a restriction endonuclease and methyl-transferase system, bacteriocin and immunity protein, and partitioning proteins. CDSB21 encoding the predicted replication protein of p22702B had 92% amino acid identity to the putative replication protein of plasmid pKA1 from *V. cholerae*. In a previous study we described the complete nucleotide sequence of plasmid p09022 (45) isolated from *Vibrio* sp. 09022 most closely-related to *V. campbellii*. Plasmids p22702B and p09022 were a similar size (28.8-kb and 31-kb; respectively), and had predicted replication proteins had 85% amino acid identity. The partitioning proteins *parA* and *parB* encoded by CDSsB14-15 and hypothetical proteins encoded by CDSsB18-21 of p22702B and were ≥84% identical at the amino acid level to proteins encoded by CDSs of p09022 and pKA1. These CDSs had conserved organization in p22702B, p09022, and pKA1, as well as direction of transcription and predicted amino acid lengths.

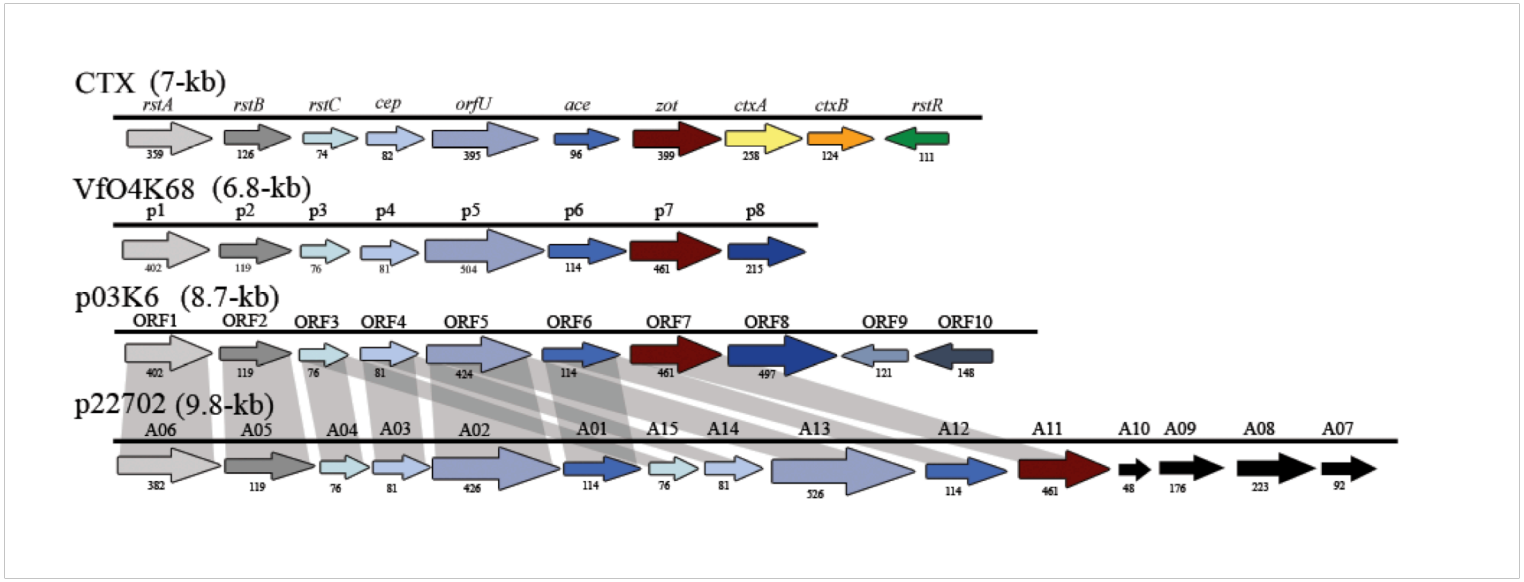


Figure 3.5. Genetic organization of the putative prophage p22702A compared to CTXϕ of *V. cholerae*, VfO4K68 of *V. parahaemolyticus* O4:K68 (20), and pO3K6, the replicative form of f237 of *V. parahaemolyticus* O3:K6 (72). The predicted amino acid sequences of a majority of the CDSs have high identity to those of f237; however, several of these conserved genes are present in duplicate copies of p22702A. Also, similar to that observed for VfO4K68, the putative virulence marker, ORF8 typical of the pandemic phage f237 (72) was not present.

Table 3.6. CDSs of the *V. parahaemolyticus* plasmid p22702A

CDS	Coordinates	Predicted aa size	Protein homolog	Phage	Organism	% Identity (Similarity)	Accession no.	e-value
A1	463-119	114	Hypothetical protein, ORF6 (114 aa)	f237	*V. parahaemolyticus*	100 (100)	NP_797936	7e-49
A2	1745-465	426	Hypothetical protein, ORF5 (504 aa)	f237	*V. parahaemolyticus*	91 (95)	NP_797935	0.0
A3	2394-2149	81	Major coat protein, ORF4 (88 aa)	f237	*V. parahaemolyticus*	96 (97)	NP_797933	1e-18
A4	2630-2400	76	Hypothetical protein, ORF3 (76 aa)	f237	*V. parahaemolyticus*	100 (100)	NP_797932	2e-33
A5	2993-2634	119	Hypothetical protein, ORF2 (119 aa)	f237	*V. parahaemolyticus*	77 (90)	NP_797931	2e-47
A6	4145-2997	382	RstA, Replication initiation factor (402 aa)	f237	*V. parahaemolyticus*	98 (99)	NP_797930	0.0
A7	4344-4129	71	Hypothetical protein VP1550 (71 aa)		*V. parahaemolyticus*	97 (97)	NP_797929	1e-28
A8	4622-4347	91	No homolog					
A9	5293-4625	222	Hypothetical protein (222 aa)		*Vibrio* sp. Ex25	94 (96)	ZP_01476538	3e-46
A10	5964-5821	47	Hypothetical protein VP1559 (47 aa)		*V. parahaemolyticus*	97 (97)	NP_797938	1e-10
A11	7349-5964	461	Zot (zonula occludens toxin), ORF7 (461 aa)	f237	*V. parahaemolyticus*	99 (99)	NP_797937	0.0
A12	7698-7354	114	Hypothetical protein, ORF6 (114 aa)	f237	*V. parahaemolyticus*	100	NP_797936	7e-49
A13	8980-7700	426	Hypothetical protein, ORF5 (504 aa)	f237	*V. parahaemolyticus*	91 (95)	NP_797935	0.0
A14	9629-9384	81	Hypothetical protein, ORF4 (81 aa)	f237	*V. parahaemolyticus*	96 (97)	NP_797933	1e-18
A15	9865-9635	76	Hypothetical protein, ORF3 (76 aa)	f237	*V. parahaemolyticus*	100	NP_797932	2e-33

Genes of interest on p22702B included a *mazG*-like protein, a bacteriocin-like protein, and an immunity protein (Figure 3.6, Table 3.7). CDSB6 encoded a putative MazG protein with 51% amino acid identity to the MazG of the *V. parahaemolyticus* RIMD2210633 genome. CDSB16 encoded a hypothetical protein with a conserved metallo-beta-lacto hydrolase domain (COG2333) typical of *comEC* proteins. The predicted protein of CDSB16 had 34% aa identity to a hypothetical protein of *Burkholderia pseudomallei* (Table 3.7).

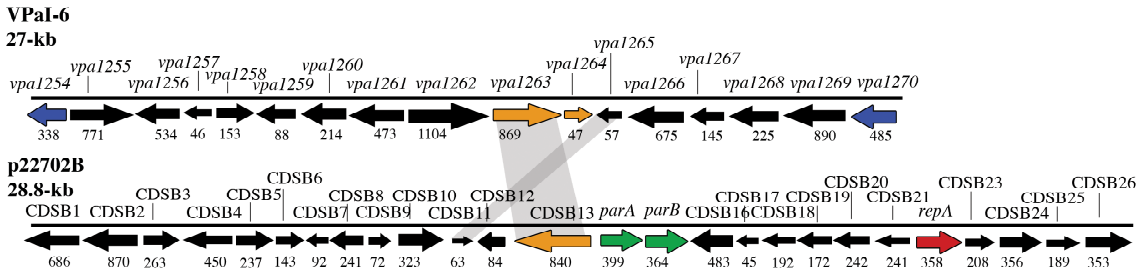


Figure 3.6. Genetic organization of *V. parahaemolyticus* plasmid p22702B (28.8-kb) compared to the genomic island VPaI-6. The approximate size of the arrows and number below each arrow indicates the amino acid sizes. The orientation of transcription is indicated by the direction of the arrows. Shaded regions show homology of p22702B CDSs to VPaI-6 genes.

Table 3.7. CDSs of the *V. parahaemolyticus* plasmid p22702B

CDS	Coordinates	Predicted aa length	Protein homolog	% Identity (Similarity)	Plasmid or Phage	Organism	Accession no.	e-value
B1	2990-930	686	Cytosine-specific DNA methyltransferase (417 aa)	62 (75)		*Escherichia coli*	CAA57629	5e-144
B2	5599-2987	870	Topoisomerase IV, subunit B (626 aa)	98 (99)		*Vibrio parahaemolyticus*	NP_796809	0.0
B3	5684-6472	262	Type II restriction endonuclease, TdeIII	25 (45)		*Chlorobium chlorochromatii*	YP_378649	4e-06
B4	7885-6530	451	Cytosine-specific DNA methyltransferase (417 aa)	62 (76)		*Escherichia coli*	CAA57629	2e-145
B5	8013-8717	234	Cytosine-specific DNA methyltransferase (417 aa)	62 (76)		*Escherichia coli*	CAA57629	2e-145
B6	8680-9111	143	mazG-related protein (129 aa)	51 (71)		*Vibrio parahaemolyticus*	NP_800286	2e-15
B7	9367-9089	92	Hypothetical protein (91 aa)	50 (72)		*Marinobacter* sp. ELB17	ZP_01739468	2e-21
B8	10202-9477	241	Hypothetical protein (241 aa)	79 (91)		*Vibrio* sp. Ex25	ZP_01473542	2e-109
B9	10294-10509	70	No homolog					
B10	10651-11619	292	Hypothetical protein VFA0519 (299 aa)	51 (68)		*Vibrio fischeri*	YP_206477	6e-71
B11	11673-11861	62	Hypothetical protein VPA1265 (57 aa)	57 (78)		*Vibrio parahaemolyticus*	NP_800774	0.17
B12	12147-11893	84	Chain A, ColE7 immunity protein (87 aa)	56 (73)		*Escherichia coli*	CAA45165	2e-18
B13	14723-12150	840	Hypothetical protein VPA1263 (869 aa)	93 (95)		*Vibrio parahaemolyticus*	NP_800773	0.0
B14	15014-16213	399	Partitioning protein, ParA (399 aa)	90 (94)	pKA1	*Vibrio cholerae*	AAW51293	0.0
B15	16213-17283	356	Partitioning protein, ParB (356 aa)	82 (92)	pKA1	*Vibrio cholerae*	AAW51292	3e-149
B16	18875-17427	482	Hypothetical protein (526 aa)	34 (48)		*Burkholderia pseudomallei*	YP_332201	3e-06
B17	19775-19197	192	Type IV pilin, putative (184 aa)	32 (53)		*Vibrio cholerae*	ZP_230505	2e-19
B18	20328-19810	172	Hypothetical protein SOA0010 (172 aa)	84 (92)	pKA1	*Vibrio cholerae*	AAW51298	3e-81
B19	20789-20325	153	Hypothetical protein SOA0008 (154 aa)	91 (94)	pKA1	*Vibrio cholerae*	AAW51297	7e-55
B20	21789-21064	240	Hypothetical protein SOA0007 (240 aa)	93 (96)	pKA1	*Vibrio cholerae*	AAW51296	2e-123
B21	22730-23803	357	Replication initiation protein (230 aa)	92 (97)	pKA1	*Vibrio cholerae*	AAW51295	3e-112
B22	24401-25027	208	Site-specific recombinase (198 aa)	86 (92)		*Vibrio cholerae*	AAX89421	1e-85
B23	25037-26104	355	Hypothetical protein (272 aa)	44 (62)		*Acidovorax* sp. JS42	YP_985143	5e-20
B24	26226-26795	189	Resolvase (179 aa)	73 (84)		*Psychromonas ingrahamii*	YP_941712	8e-66
B25	26815-27866	349	HNH endonuclease (242 aa)	51 (62)		*Burkholderia multivorans*	ZP_01567608	5e-50
B26	27891-28346	151	No homolog					

CDSB13 of p22702B is identical to *vpa1263* of the pandemic genomic island VPaI-6.

We identified a CDS of p22702B with a predicted protein that had significant amino acid identity (93%) to a hypothetical protein of the predicted *V. parahaemolyticus* RIMD2210633 genomic island VPaI-6 (Figure 3.6). The predicted proteins of CDSB13 and *vpa1263* contained a 180 amino acid functional domain characteristic of S-type pyocins (pfam06958) indicating these proteins may function as nuclease-type bacteriocins (Figure 3.7). The N-terminal region of the p22702B bacteriocin had high identity to a LysM domain (smart00257) that was shared with *vpa1263,* a hypothetical protein of *V. harveyi*, and a hypothetical protein of *V. cholerae* (Figure 3.7). A putative bacteriocin immunity protein CDSB12 with 56% aa identity to a ColE7 immunity protein is encoded upstream of the bacteriocin (CDSB13) (Figure 3.6, Table 3.7). Sequence comparison of CDSB12 encoding the predicted immunity protein to the entire VPaI-6 genomic island indicated this gene is unique to p22702B.

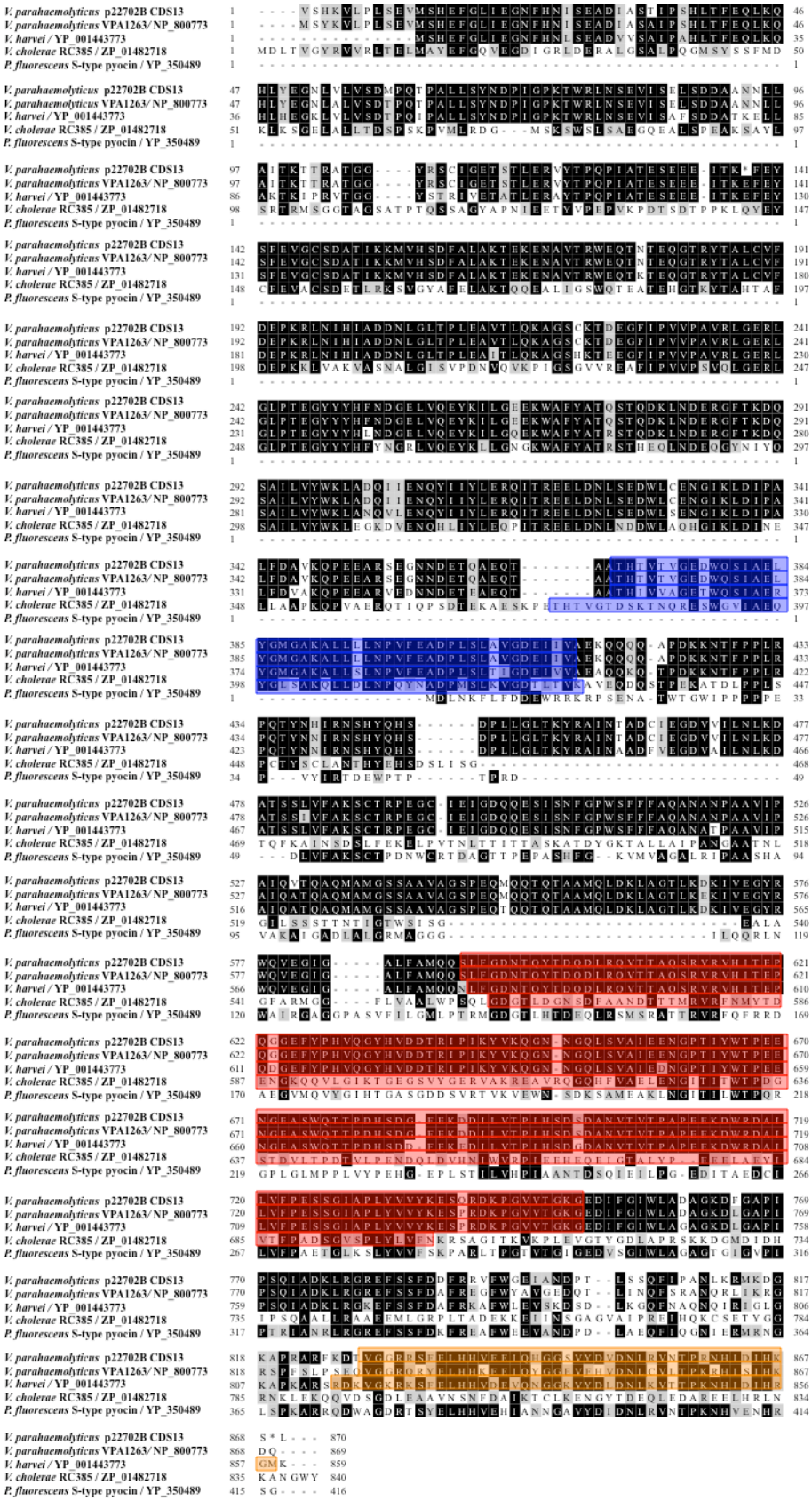

V. parahaemolyticus p22702B CDS13
V. parahaemolyticus VPA1263/NP_800773
V. harvei / YP_001443773
V. cholerae RC385 / ZP_01482718
P. fluorescens S-type pyocin / YP_350489

Figure 3.7. Amino acid sequence alignment of the predicted protein encoded by CDS13 of p22702B compared to VPA1263 of *V. parahaemolyticus* RIMD2210633, an S-type pyocin of *P. aeruginosa*, a hypothetical protein of *V. harveyi,* and a hypothetical protein of *V. cholerae*. The blue shaded region indicates a LysM binding domain, red shows a S-type pyocin domain, and orange indicates an HNH nuclease domain.

Distribution of MGEs among *V. parahaemolyticus* clinical and environmental strains.

Presumptive *V. parahaemolyticus* strains were isolated from environmental samples by plating on TCBS and green colonies screened for genetic elements. Any strain that had an extrachromosomal element was determined to be *V. parahaemolyticus* by PCR screening for the thermolabile hemolysin (*tl),* which has been previously used to identify *V. parahaemolyticus* strains (65). A modified Kieser acid phenol extraction method was used to separate supercoiled extrachromosomal elements from chromosomal DNA to examine the occurrence of extrachromosomal elements among *V. parahaemolyticus* clinical and environmental strains. In chapter 2, we demonstrated that of 63 *V. parahaemolyticus* clinical strains and 422 *Vibrio* environmental strains screened there were 18 (29%) clinical and 157 (37%) environmental strains that had one or more extrachromosomal elements identified using the Kieser extraction method. The Kieser method can be used to reproducibly isolate elements as small as 2-kb or as large as 300-kb (54). All of the MGE-bearing *V. parahaemolyticus* clinical strains and a subset of the *Vibrio* environmental strains are listed in Table 3.8 with the number of extrachromosomal bands that were reproducibly identified from at least three or more independent cultures. Of the *V. parahaemolyticus* clinical strains that had extrachromosomal elements approximately 75% of these strains had elements that were ≤23-kb .

Diversity and distribution of a *V. parahaemolyticus* plasmid family. To determine the diversity and distribution of a conserved *Vibrio* replication protein gene *rep* previously characterized from the *V. campbellii* plasmid p09022 and the *V. parahaemolyticus* plasmid p22702B, we designed primers to screen and sequence the replication protein-encoding genes. Of 41 *Vibrio* environmental and 12 *V. parahaemolyticus* clinical strains examined that were identified by Kieser screening to have extrachromosomal elements there were 14 environmental strains that tested positive by a PCR screen for the conserved plasmid *rep* gene (Table 3.8).

Table 3.8. MGEs of *V. parahaemolyticus* clinical strains and *Vibrio* environmental strains

						No. Extrachromosomal Bands					
Strain Id	Year	Location	Serotype	Source	*tl*	<23-kb	23-60-kb	>60-kb	*repA*	*rstA*	ORF8
RIMD2210633	1996	Japan	O3:K6	clinical	+	1	0	0	-	+	+
K1533	2004	CT	O3:K6	clinical	+	0	0	0	ND	+	+
K1223	2004	CO	O3:K6	clinical	+	1	1	0	-	+	+
17802	1965	Japan	O3:K6	clinical	+	0	0	0	ND	-	-
K1378	2004	AZ	O3:K6	clinical	+	0	0	0	ND	+	+
K1325	2004	UT	O3:K6	clinical	+	0	0	0	ND	+	+
F5847	1998	TX	O3:K6	clinical	+	1	0	0	-	+	+
F8949	2002	MA	O3:K6	clinical	+	0	0	0	ND	+	+
F9083	2002	AZ	O3:K6	clinical	+	0	0	0	ND	+	+
K3584	2006	NY	O3:K6	clinical	+	4	0	0	-	+	+
K4377	2006	MD	O3:K6	clinical	+	0	0	0	ND	+	+
K4435	2006	GA	O3:K6	clinical	+	0	0	0	ND	+	+
F8186	2001	NY	O4:K12	clinical	+	0	0	0	ND	-	-
F8023	2001	GA	O4:K12	clinical	+	0	0	0	ND	-	-
K0851	2004	LA	O1:Kuk	clinical	+	0	0	0	ND	+	-
K1461	2004	MA	O4:K12	clinical	+	0	0	1	-	-	-
K0081	2003	AZ	O6:K18	clinical	+	0	0	0	ND	-	-
K0850	2004	LA	O1:K25	clinical	+	0	0	0	ND	-	-
F7979	2001	MN	O5:K17	clinical	+	1	0	0	-	+	+
F5052	1997	WA	O4:K12	clinical	+	0	0	0	ND	+	+
F6179	1998	CT	O6:K18	clinical	+	1	0	0	-	+	+
F7044	2000	TX	O4:K12	clinical	+	0	0	0	ND	+	+
F8937	2002	NY	O4:K10	clinical	+	1	0	0	-	-	-
K4237	2006	NY	O10:K56	clinical	+	1	0	0	-	+	-
K4250	2006	NY	O4:K63	clinical	+	0	0	1	-	+	-
K4376	2006	MD	O8:K21	clinical	+	1	0	0	-	+	-
K4434	2006	MS	O6:K18	clinical	+	0	0	0	ND	+	+
K4638	2007	NY	O10:Kuk	clinical	+	0	0	0	ND	+	+
K4763	2007	VA	O5:Kuk	clinical	+	1	0	0	-	+	-
K4981	2007	OK	O1:Kuk	clinical	+	0	0	0	ND	+	+
K5067	2007	SD	O1:K56	clinical	+	0	0	1	-	-	-
K5281	2007	WA	O4:K12	clinical	+	0	0	0	ND	-	-
K5323 g	2007	VA	O5:K17	clinical	+	0	0	0	ND	+	+
K5330	2007	TX	O5:Kuk	clinical	+	0	0	0	ND	+	-
22702	1998	GA	O5:Kuk	sediment	+	1	1	0	+	+	-
SG63	2006	GA	O10:Kuk	water	+	1	0	0	-	+	-
SG167	2006	GA	ND	water	+	0	1	0	+	-	-
SG176	2006	GA	O5:Kuk	water	+	0	0	1	-	+	-
SG189	2006	GA	ND	water	-	0	0	2	+	-	-
SG205	2006	GA	ND	sediment	+	1	0	0	-	-	-
SG209	2006	GA	O2:K28	sediment	+	0	1	0	-	-	-
SG212	2006	GA	ND	water	+	1	0	0	-	-	-
SG214	2006	GA	O1:K9	water	+	1	0	1	-	-	-
SG219	2006	GA	ND	water	-	0	1	1	+	-	-
SG244	2006	GA	O6:Kuk	water	+	0	1	1	-	+	+
SG257	2006	GA	O8:Kuk	sediment	+	0	0	1	-	-	-
SG258	2006	GA	O1:Kuk	sediment	+	0	2	0	-	-	-
SG259	2006	GA	ND	sediment	+	0	0	1	-	+	+
SG285	2006	GA	O10:Kuk	sediment	+	0	1	0	-	-	-
SG328	2006	GA	ND	water	+	0	0	1	-	+	-
SG354	2006	GA	O2:K28	water	+	0	1	0	-	+	+
SG364	2006	GA	O1:Kuk	sediment	+	0	0	1	+	-	-
SG372	2006	GA	O5:Kuk	sediment	+	0	1	0	-	-	-
SG395	2006	GA	ND	water	-	0	0	1	-	-	-
SG401	2006	GA	O5:K61	sediment	+	0	1	0	-	-	-
SG409	2006	GA	O1:K33	sediment	+	0	1	0	-	-	-
SG417	2006	GA	O4:K67	sediment	+	0	0	3	-	-	-
AF2	2006	FL	O4:K9	oyster	+	0	1	0	+	-	-
AF5	2006	FL	O4:K9	oyster	+	0	1	0	+	+	+
AF23	2006	FL	ND	sediment	-	0	0	1	+	-	-
AF34	2006	FL	O5:K37	sediment	+	0	0	2	-	-	-
AF46	2006	FL	O11:Kuk	water	+	0	0	2	-	-	-
AF47	2006	FL	ND	water	+	0	0	2	-	-	-
AF56	2006	FL	O6:K37	sediment	+	0	0	1	-	-	-
AF64	2006	FL	ND	sediment	+	0	1	0	-	-	-
AF67	2006	FL	O5:K37	sediment	+	0	0	2	-	+	+
AF91	2006	FL	ND	sediment	+	2	0	0	-	-	-
J-C2-15	1998	NC	O1:Kuk	sediment	+	0	0	1	+	-	-
J-C2-34	1998	NC	O5:K19	sediment	+	0	0	1	-	+	-
J-M1-23	1998	NC	O1:K32	sediment	+	0	0	1	+	-	-
S-M2-3-B3	1998	NC	O5:K17	sediment	+	1	0	1	-	-	-
7-SG81	2007	GA	ND	sediment	+	0	0	1	+	+	-
7-SG165	2007	GA	ND	sediment	+	0	0	1	+	-	-
7-SG204	2007	GA	ND	sediment	+	0	0	1	+	+	-
7-SG291	2007	GA	ND	water	-	0	1	0	+	-	-
7-SG326	2007	GA	ND	water	-	0	1	0	+	-	-

Phylogenetic analysis of *rep* genes detected among *Vibrio* environmental strains was conducted compared to phylogenetic analysis of concatenated housekeeping genes *recA* and *gyrB* to examine the evolution of plasmids associated with *V. parahaemolyticus* and closely-related environmental strains (Figure 3.8). The nucleotide sequences of the *rep* genes exhibited 71-74% identity to CDSB21 of p22702B while the amino acid sequences were 91-93% identical to that encoded by CDSB21. The *rep* sequences obtained from *V. parahaemolyticus* strains with extrachromosomal bands ≥60-kb formed a group that had little branching. In contrast, the *rep* sequences obtained from *V. campbellii* and *V. harveyi* hosts exhibited deeper branches (Figure 3.8).

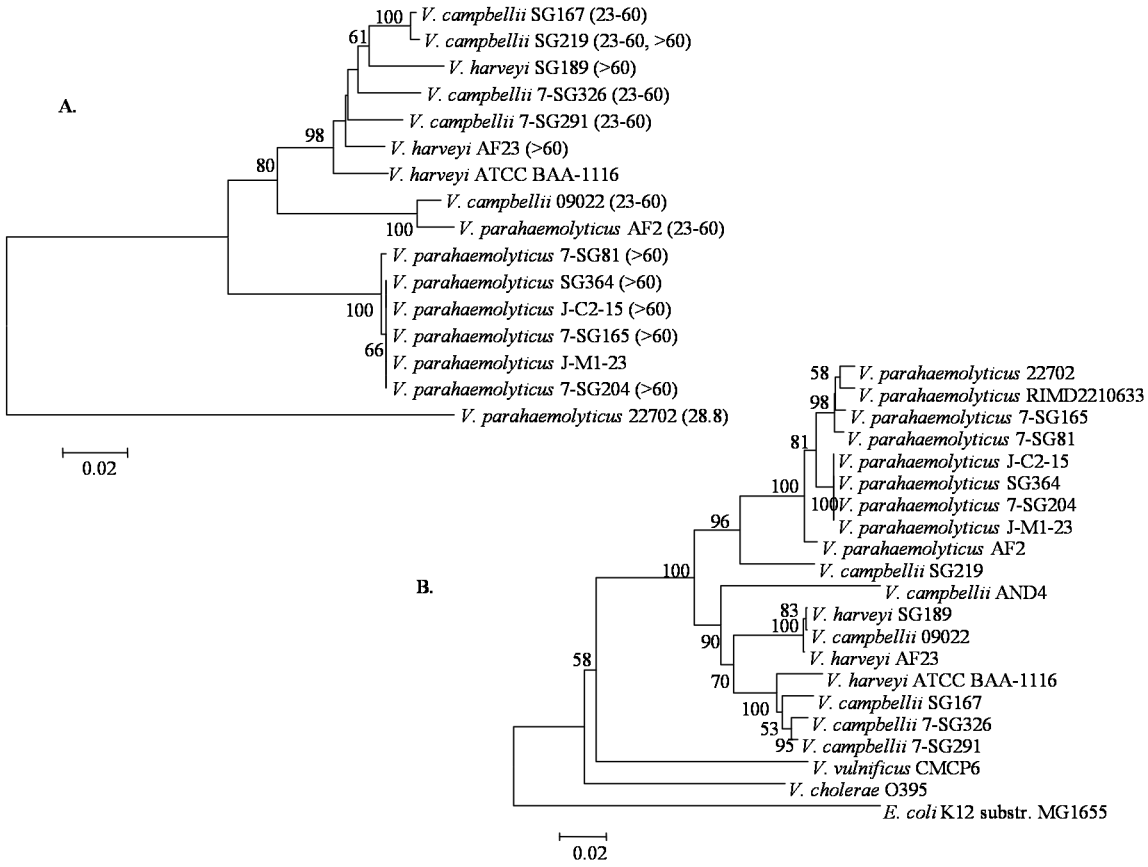


Figure 3.8. Neighbor-joining tree of genomic *recA-gyrB* concatenated nucleotide sequences from plasmid-bearing *V. parahaemolyticus* (n= 8), *V. harveyi* (n= 3), and *V. campbellii* (n= 4) environmental strains (A). Phylogenetic analysis of nucleotide sequences of the plasmid replication proteins from plasmid-bearing *V. parahaemolyticus, V. harveyi,* and *V. campbellii* environmental strains (B). The trees were constructed using the Kimura 2-parameter model and values represent 1,000 bootstrap replications. Only values ≥50 are shown. The scale bar represents 0.02 nucleotide substitutions per site.

The presence of the sequenced *rep* genes on the plasmids was confirmed by Southern hybridization of a *rep* probe generated from 22702 DNA to Kieser-extracted plasmid DNA that had been transferred to a positively charged membrane (Figure 3.9). The *rep* probe strongly hybridized to the supercoiled plasmid DNA and not to the linearized chromosomal DNA from any of the selected plasmid-bearing strains shown (Figure 3.9). In addition, the *rep* probe did not hybridize to chromosomal DNA of the plasmid-free *V. parahaemolyticus* RIMD2210633 (Figure 3.9). The selected plasmids that were positive for the 22702 rep-type were all isolated from *Vibrio* environmental strains. Most of the plasmids that had the 22702 rep-type were 23-60-kb; however, there were select plasmids that were ≥60-kb (Table 3.8, Figure 3.9).

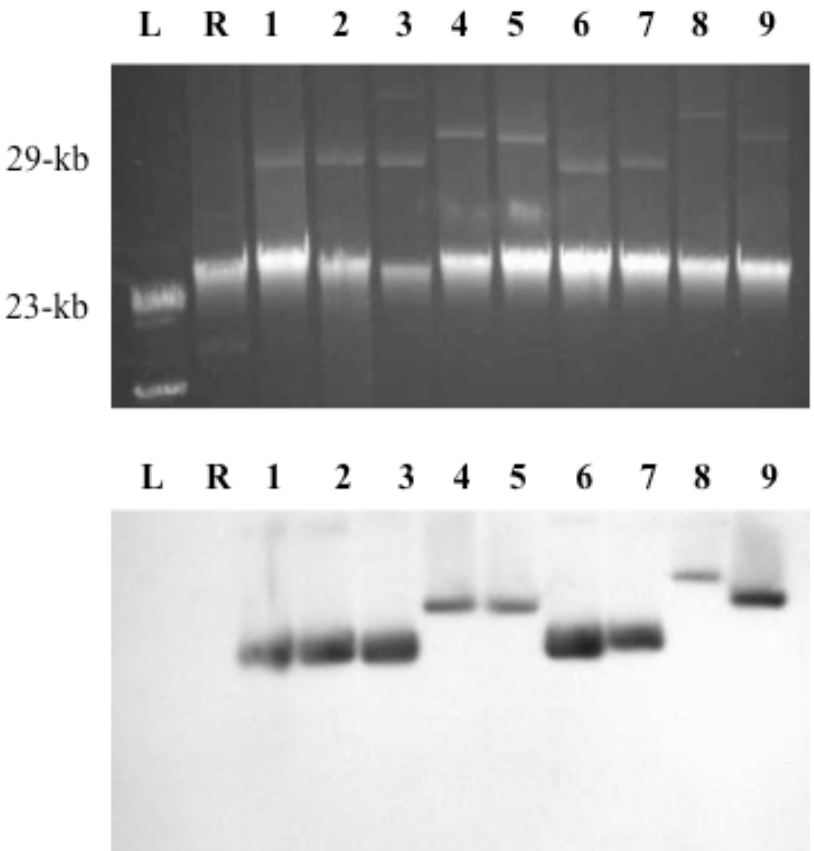


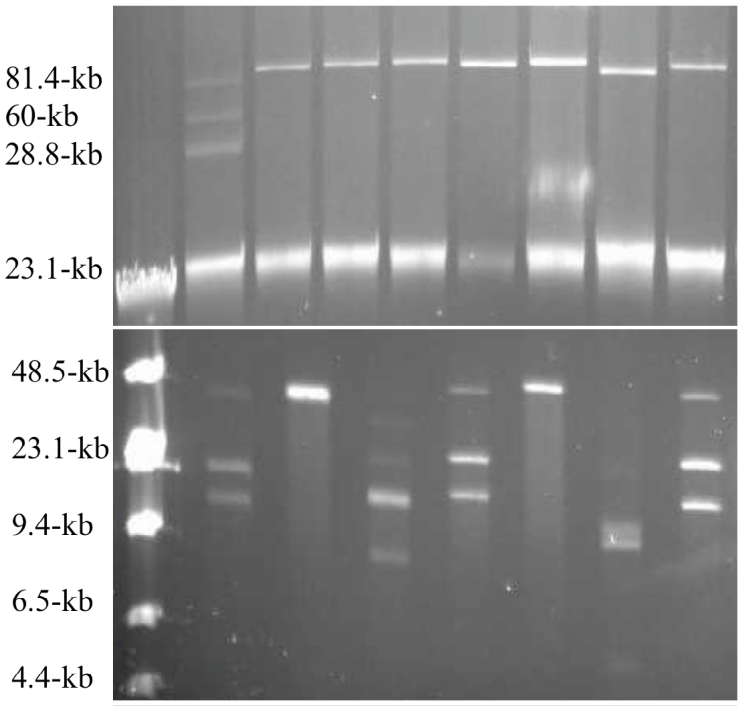
Figure 3.9. Southern hybridization of a *V. parahaemolyticus* 22702 *rep* probe to plasmid DNA of *V. parahaemolyticus, V. harveyi,* and *V. campbellii* environmental strains. Lanes 1 to 9: plasmid DNA from *V. parahaemolyticus* 22702, *V. campbellii* 09022, *V. campbellii* SG219, *V. parahaemolyticus* SG364, *V. parahaemolyticus* J-M1-23, *V. parahaemolyticus* 7-SG291, *V. parahaemolyticus* AF2, *V. parahaemolytiucs* 7-SG81, and *V. harveyi* AF23. Lambda Hind III ladder, lane L; *V. parahaemolyticus* RIMD2210633 chromosomal DNA, lane R. Chromosomal DNA is present in each sample as a sheared band at approximately 23-kb.

Phylogenetic analysis of a concatenation of the *recA* and *gyrB* nucleotide sequences PCR amplified from genomic DNA of the host strains was used to identify select MGE-bearing environmental host strains as *V. parahaemolyticus, V. harveyi,* or *V. campbellii* (Figure 3.9). The *recA-gyrB* tree showed a distinct separation of the *V. parahaemolyticus* strains from the *V. campbellii* and *V. harveyi* host strains (Figure 3.8). *V. parahaemolyticus* environmental strains J-C2-15, SG364, 7-SG204, and J-M1-23 were isolated from geographically separated samples from different years (Table 3.8); however, these strains formed a group in the housekeeping gene phylogenetic analysis (Figure 3.8).

The structural diversity of select plasmids was examined by restriction endonuclease digestion of gel-purified plasmid DNA. Of the *V. parahaemolyticus* strains identified with MGEs, there were seven that had bands of approximately 86-kb (Figure 3.10). Restriction digestion of these similarly-sized elements resulted in distinct four profiles. The restriction profile of the element isolated from *V. parahaemolyticus* clinical

strain K1461 was similar to that from two *V. parahaemolyticus* environmental strains SG176 and J-C2-34 (Figure 3.10) from GA and NC, respectively. These elements were digested into three fragments that were approximately 48-kb, 23-kb, and 15-kb (Figure 3.10). The restriction profile of the element from the *V. parahaemolyticus* clinical strain K4250 was four bands of approximately 40-kb, 23-kb, 15-kb, and 8-kb. The elements from the *V. parahaemolyticus* clinical strain K5067 and the environmental strain SG214 were digested into single 48-kb bands (Figure 3.10). The element from *V. parahaemolyticus* SG395 was digested bands that were approximately 23-kb, multiple 9-12-kb bands, and a 5-kb band. A group of large plasmids (>60-kb) isolated from *V. parahaemolyticus* environmental strains that had similar *rep* genes (Figure 3.9) were digested to form two different profiles (Figure 3.10). Four of the plasmids from the *V. parahaemolyticus* environmental strains that had conserved *rep* genes were restricted into three bands of approximately 40-kb, 23-kb, 8-kb (Figure 3.10). An additional plasmid isolated from *V. parahaemolyticus* environmental strain 7-SG81 was digested into three bands that were approximately 65-kb, 48-kb, and 8-kb (Figure 3.10).

A.


81.4-kb
60-kb
28.8-kb
23.1-kb
48.5-kb
23.1-kb
9.4-kb
6.5-kb
4.4-kb

B.

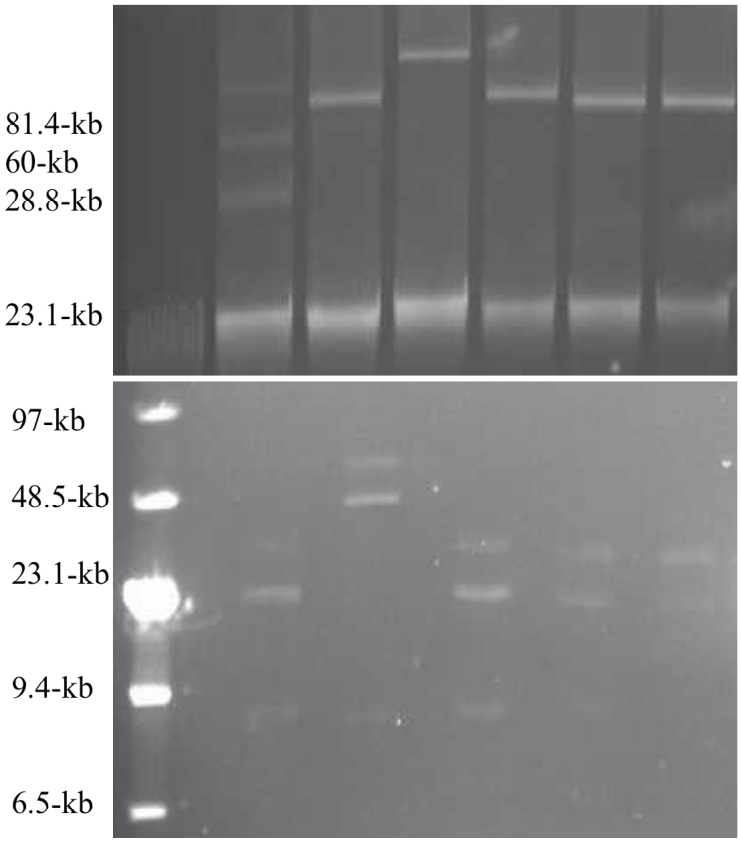

81.4-kb
60-kb
28.8-kb
23.1-kb
97-kb
48.5-kb
23.1-kb
9.4-kb
6.5-kb

Figure 3.10. Restriction digestion patterns of select *Vibrio* extrachromosomal elements: (A) an approximately 84-kb element present in *V. parahaemolyticus* clinical and environmental strains. Lane 1, plasmids of known sizes (81.4-kb, 60-kb, 28.8-kb); Lanes 2-4, *V. parahaemolyticus* clinical strains K1461, K5067, K4250, and Lanes 5-8, *V. parahaemolyticus* environmental strains SG176, SG214, SG395, and J-C2-34. (B) Plasmids isolated from *V. parahaemolyticus* environmental strains (left to right: SG364, 7-SG81, 7-SG204, J-C2-15, and J-M1-23) that have *rep* genes with significant nucleotide identity (≥ 99%). The top images in (A) and (B) show supercoiled plasmid DNA and linearized chromosomal DNA at 23-kb. The bottom images in each set show PacI and SbfI restriction digested plasmid profiles.

Diversity and distribution of *V. parahaemolyticus* filamentous phage. A total of 43 *Vibrio* environmental and 33 *V. parahaemolyticus* clinical strains were PCR screened for the phage replication gene *rstA*. The *rstA* screen included strains that did not have detectable extrachromosomal bands since filamentous phage may be present in the chromosome as a co-integrant and therefore would not be detected by the Kieser extraction method. Of the *Vibrio* environmental strains screened for *rstA,* 28% (12/43) were positive whereas 73% (24/33) of the *V. parahaemolyticus* clinical strains examined were positive for *rstA*.

Sequence and phylogenetic analysis of *rstA* nucleotide sequences obtained from *V. parahaemolyticus* clinical and environmental strains revealed these sequences had 95-100% nucleotide identity to the *rstA* of the filamentous phage f237. Of the 36 *V. parahaemolyticus* strains that tested positive for *rstA* there were 20 strains that had at

least one detectable extrachromosomal band ≤23-kb and 23 with at least one extrachromosomal band >23-kb (Table 3.8). There were two strains that had extrachromosomal bands >23-kb and ≤23-kb. In addition, there were 15 *V. parahaemolyticus* clinical strains examined that had no detectable extrachromosomal bands. A neighbor-joining tree constructed with select *V. parahaemolyticus* clinical and environmental strain *rstA* nucleotide sequences showed these sequences are highly conserved (≥90% nucleotide identity) and there are some sequences that are more related than others (Figure 3.11). For example, the *rstA* sequences associated with the O3:K6 strains RIMD2210633 and K1223 were related to the *rstA* of the sequenced filamentous vibriophage vfO4K68 (Figure 3.11). Also, the *rstA* sequences of the *V. parahaemolyticus* clinical strains K5323g and K4376 are closely-related; however, these strains have different serotypes (Figure 3.11). There was no detectable pattern of vibriophage association with strains belonging to certain *V. parahaemolyticus* serotypes. In addition, *rstA* gene sequences amplified from *V. parahaemolyticus* clinical and environmental strains in housekeeping gene phylogenies (43).

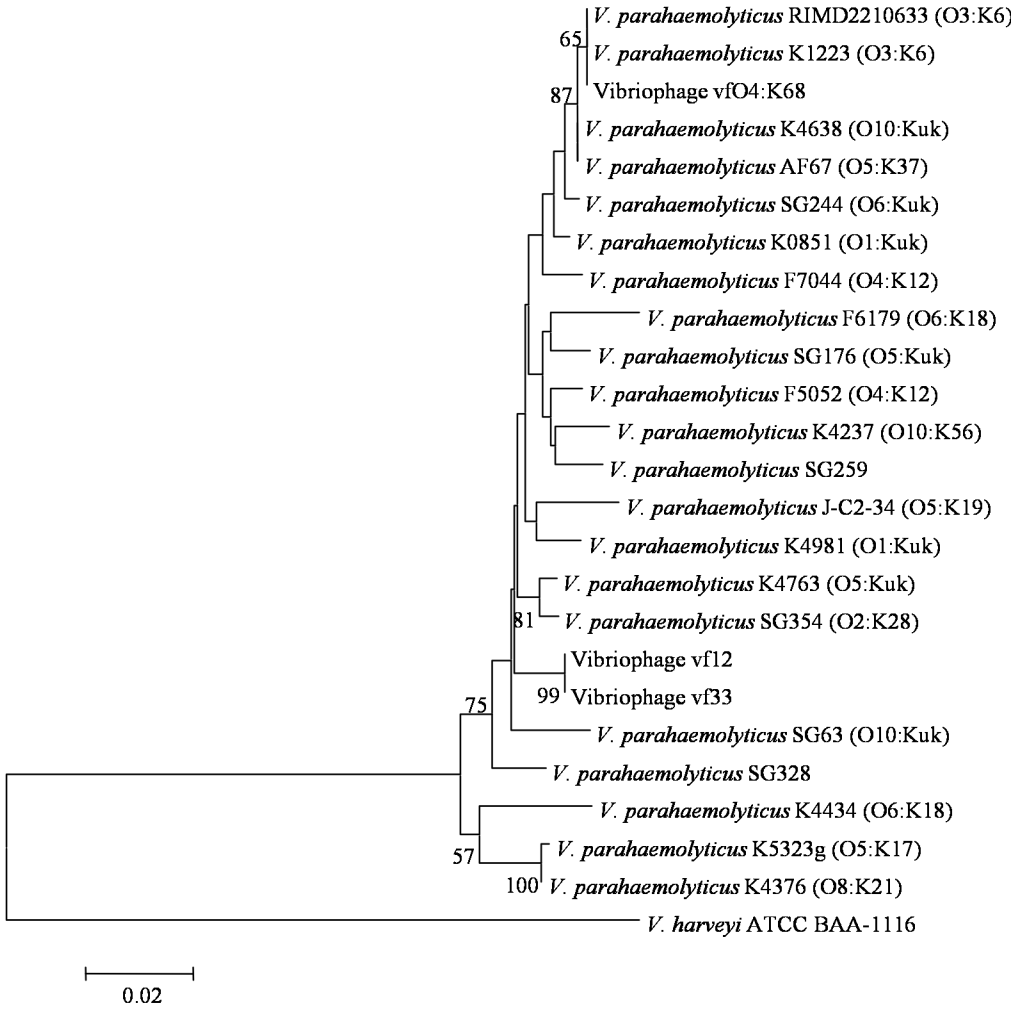


Figure 3.11. Neighbor-joining tree of *rstA* nucleotide sequences from total DNA of *V. parahaemolyticus* clinical (n= 12) and environmental (n= 8) strains. The tree was constructed using the Kimura 2-parameter model with 1,000 bootstrap replications. Only values ≥50 are shown. The scale bar represents 0.02 nucleotide substitutions per site.

Plasmid- and chromosomally-encoded bacteriocin *vpa1263*. The frequency of the predicted bacteriocin *vpa1263* was analyzed among *V. parahaemolyticus* clinical (n= 33) and environmental (n= 38) strains. In addition, the frequency of another VPaI-6 gene, *vpa1260,* that did not occur on p22702B was analyzed. Both *vpa1260* and *vpa1263* were detected among 11 *V. parahaemolyticus* clinical strains and none of the environmental strains examined. There were 10 *V. parahaemolyticus* strains that had both *vpa1260* and *vpa1263* that were detected among the *V. parahaemolyticus* clinical strains examined (Table 3.8).

To investigate the function of the *vpa1263* gene, we constructed a Δ*vpa1263* strain of the *V. parahaemolyticus* O3:K6 pandemic strain RIMD2210633. *V. parahaemolyticus* strains were screened for their ability to inhibit growth of *V. parahaemolyticus* and diverse *Vibrio* spp. using a bacteriocin screening assays that were previously used to demonstrate the production of a bacteriocin-like substance by *V. harveyi* (63, 85). We did not detect the production of an inhibitory substance using these previously developed methods (data not shown). In addition, we used a spot-on-lawn assay but did not detect any inhibition by any of the *V. parahaemolyticus* strains examined or by *V. harveyi* ATCC BAA-1116, which also possesses a homolog of the *vpa1263* bacteriocin (data not shown). Reverse transcription PCR analysis (RT-PCR) showed that the S-type pyocin domain of *vpa1263* of *V. parahaemolyticus* RIMD2210633 is expressed following growth overnight in rich media at 30˚C with shaking (Figure 3.12). In addition, *vpa1263* was expressed in Δ*rpoS* and Δ*opaR* mutants of RIMD indicating that the expression of *vpa1263* is not regulated by the induction of stationary phase survival or by quorum sensing of *V. parahaemolyticus.* The *vpa1263*

pyocin domain of CDSB13 on p22702B of *V. parahaemolyticus* 22702 was not expressed.

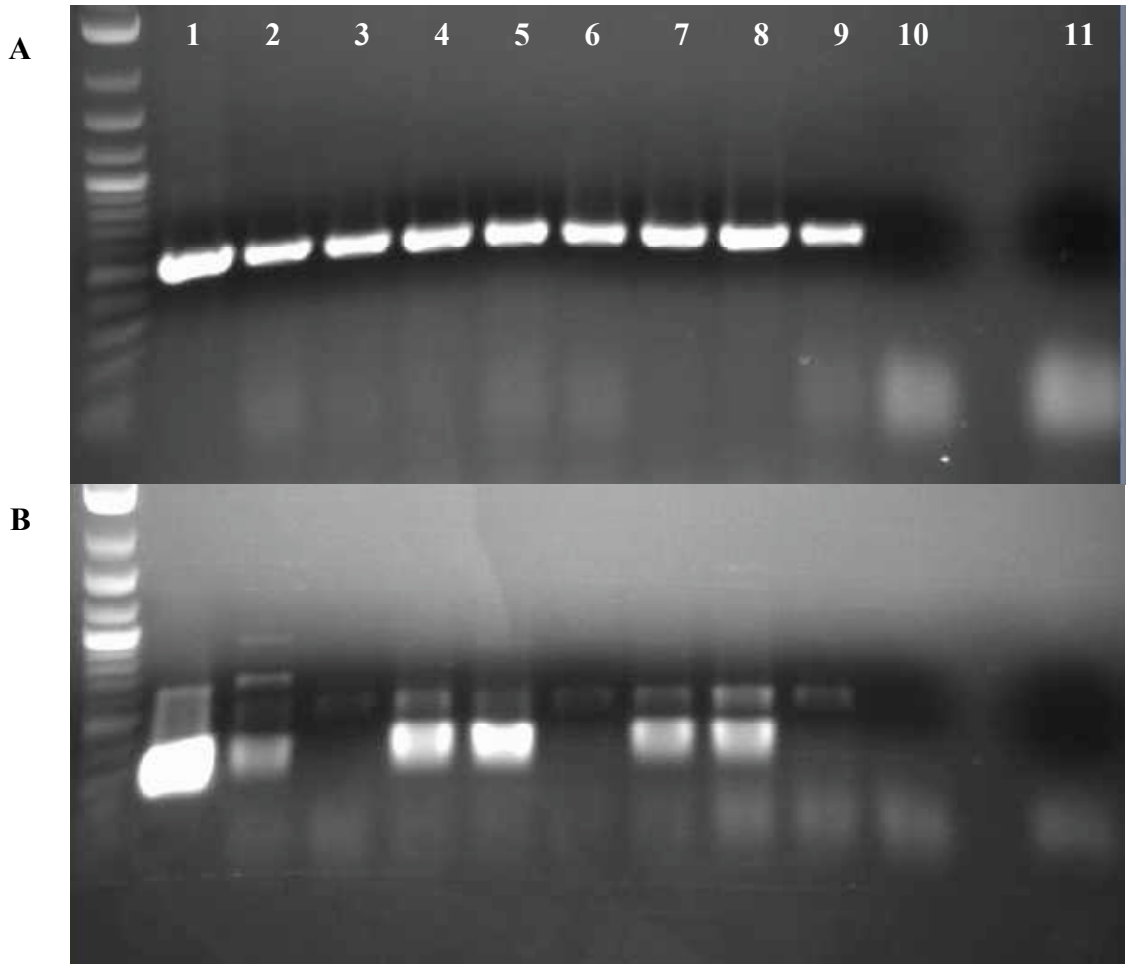


Figure 3.12. RT-PCR expression analysis of *rpoB* (A) and *vpa1263* (B) from *V. parahaemolyticus* RIMD2210633 DNA control, 1; and cDNA: *V. harveyi* ATCC BAA-1116, 2; *V. parahaemolyticus* 22702, 3; *V. parahaemolyticus* RIMD, 4; RIMD str^r, 5; Δ*vpa1263*, 6; Δ*opaR,* 7; Δ*rpoS,* 8; ATCC 17802, 9; RT no template control, 10; and PCR no template control, 11.

To determine whether the expression of *vpa1263* provides a fitness advantage for competition during logarithmic growth, we conducted a liquid competition assay as previously performed for *V. harveyi* (48). The deletion strain Δ*vpa1263* exhibited a lower cell density after 10 and 24 hours growth during competition than the wild-type competitor *V. parahaemolyticus* RIMD2210633 (Figure 3.13). There was no decrease in the detectable CFU/ml of the Δ*vpa1263* that was grown without competition.

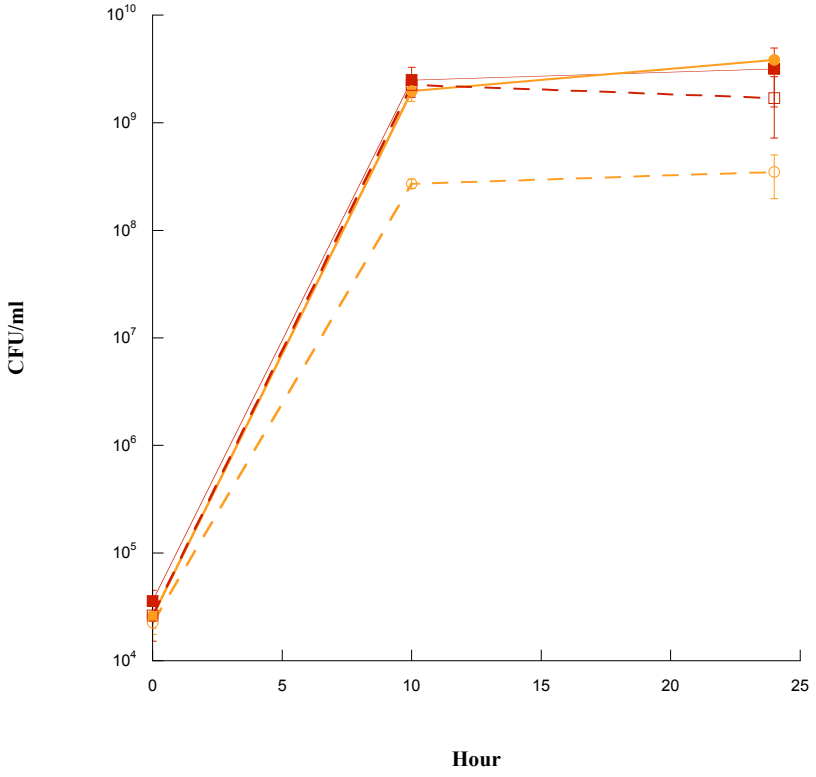


Figure 3.13. Competition of *V. parahaemolyticus* RIMD2210633 and the Δ*vpa1263* mutant strain grown in LB for 24 h with shaking. RIMD is indicated by squares while Δ*vpa1263* is indicated by circles. The solid line indicates strains that were grown without competition and the dashed line indicates strains that were grown in competition.

Discussion

V. parahaemolyticus strains have been shown to have considerable diversity (23, 40, 106) and with a high frequency of plasmids (85); however, gene content of *V. parahaemolyticus* plasmids as well as their relatedness to plasmids of closely-related *Vibrios* within the same environmental niche has not been examined. Few studies to date have examined the frequency of plasmids associated with *V. parahaemolyticus* clinical or environmental strains (19, 111). In addition, there are few sequences of plasmids or phage isolated from *V. parahaemolyticus* that are currently available in GenBank. The majority of the genetic elements that have been characterized from *V. parahaemolyticus* are filamentous phage (20, 21, 72) or small plasmids such as pSA19 (GenBank). In a previous study, we analyzed the sequence diversity of three plasmids isolated from environmental *Vibrio* spp. and demonstrated these genetic elements had considerable diversity (45). In the current study, we characterized the sequences of a plasmid and a prophage isolated from a *V. parahaemolyticus* environmental strain. In addition, we describe that the sequenced *V. parahaemolyticus* plasmid, p22702B, belongs to a plasmid family distributed among *V. parahaemolyticus, V. campbellii,* and *V. harveyi*.

In the present study, we show that plasmids isolated from *V. parahaemolyticus* environmental strains share a conserved replication protein-encoding gene with plasmids from *V. harveyi* and *V. campbellii* strains. The conserved replication protein-encoding genes were detected in environmental *Vibrios* isolated from geographically and temporally isolated samples indicating these plasmids and their host strains may move frequently. Alternately, these plasmids may be ancestrally-related and may have disseminated among related environmental *Vibrios* early on in the speciation of these

strains. The significant diversity of the structural profiles and gene content demonstrated for these plasmids supports that they are ancestral to *V. parahaemolyticus* and the closely-related *V. harveyi* and *V. campbellii* and the *rep* genes are under strong selective constraint preventing nucleotide diversification.

In addition, the demonstration of structurally-conserved plasmids present among a *V. parahaemolyticus* clinical strain and two *V. parahaemolyticus* environmental strains indicates these elements may facilitate HGT among clinical and environmental strains. The plasmid-bearing clinical strain is an O4:K12 clonal strain that possesses both *tdh* and *trh*. Previous studies have shown that pathogenicity of the O3:K6 clonal strains involves T3SS2α and *tdh* that are both present within the VPaI-7 genomic island (81). In contrast, other strains have been identified that have *trh* and T3SS2β that lack VPaI-7 (78). The clonal O4:K12 strains are frequently associated with disease outbreaks in the United States (2); however, the pathogenic mechanism of these strains is unknown. Sequence analysis of the plasmid isolated from the clinical strain and one of the plasmids isolated from an environmental strain is underway and may yield insight into additional genes involved in pathogenicity or adaptation of the O4:K12 strains.

The similarity of p22702A proteins to proteins of the filamentous vibriophage f237 and VfO4K68 of *V. parahaemolyticus* in addition to previous biochemical and phylogenetic analysis of housekeeping genes (29) further supports the identification of the environmental strain 22702 as *V. parahaemolyticus*. The similarity of p22702A to an f237 co-integrate in the *V. parahaemolyticus* RIMD2210633 genome as well as identification of the chromosomal flanking genes *vpa1550* and *vpa1559* provides evidence that *V. parahaemolyticus* 22702 may have be infected by a filamentous phage

from a pandemic strain. This may have occurred when the strain came into contact with an O3:K6 strain in a host organism or in the environment. In addition, we have shown in a previous study that *V. parahaemolyticus* 22702 fixes nitrogen and possesses a *nifH* homolog (29) indicating it was likely exposed to conditions promoting gene transfer and niche expansion. The conserved regions identified on p22702B, p09022, and pKA1 of a *V. cholerae* environmental isolate, suggests the potential for plasmid-mediated gene exchange between strains of the same species as well as of different species.

Several interesting proteins encoded by CDSs of p22702B include the MazEF addiction regulator MazG, a putative bacteriocin and immunity protein, and a competence-like DNA binding protein. MazG of other bacteria has been shown to have pyrophosphorohydrolase activity and was shown to lower intracellular levels of ppGpp (41). This indicated MazG may be involved in regulating the *mazEF* toxin-antitoxin addiction system that prevents induction of the stringent response pathway initiated during amino acid starvation (41). We show that CDSB13 has a domain characteristic of S-type pyocins and a HNH nuclease domain of nuclease-type colicins. In addition, CDSB13 and *vpa1263* possess a LysM domain that is involved in attachment to peptidoglycan (17). LysM domains are typical of cell-surface proteins of *Lactobacillus fermentum* (110) and *Agrobacterium tumefaciens* (58) and have been shown to facilitate cell-surface recognition. LysM has been identified as a widely-distributed protein motif that facilitates protein binding to peptidoglycan (17). Autolysins (N-acetylmuramoyl-L-alanine amidase) are proteins that have LysM motifs and are involved in peptidoglycan cleavage (17). Mutation of a *V. cholerae* autolysin, *amiB*, resulted in loss of swarming

motility and a switch from the rugose colony phenotype of virulent strains to a smooth colony phenotype (87).

V. harveyi has previously been shown to produce a bacteriocin-like substance that is associated with the presence of plasmid DNA and confers a competitive advantage to the host strain (48, 63). In this study, we examined the production of inhibitory-like substances by *V. harveyi* ATCC BAA-1116, *V. parahaemolyticus* RIMD2210633, and *V. parahaemolyticus* 22702 that all possess the putative bacteriocin-encoding gene. There was no detectable inhibition using multiple media types and screening methods by any of the bacteriocin-encoding strains toward *V. harveyi, V. parahaemolyticus, V. campbellii, V. fluvialis, V. fischeri,* or *E. coli.* Expression analysis of the S-type pyocin domain of *vpa1263* showed that transcripts of this region are present during logarithmic growth in rich media. We did not detect the production of an inhibitory substance on solid media by any of the strains examined in this study; however, there was evidence that strains with the *vpa1263* gene had a competitive advantage during growth in liquid. Competition of the Δ*vpa1263* strain with the isogenic wild-type strain indicated strains possessing the gene have a fitness advantage during growth in rich media. Similarly, a previous study demonstrated that a *V. harveyi* that possessed a bacteriocin-encoding plasmid had a fitness advantage over an isogenic plasmid-free strain when grown under non-limiting conditions (48). Further investigation is required to determine the contribution of the plasmid and chromosomally-encoded bacteriocins to host fitness during competition in the environment or in the human gut.

In conclusion, plasmids associated with *V. parahaemolyticus* are genetically diverse and serve as agents of gene transfer among clinical and environmental strains.

This is the first study demonstrating the presence of a conserved plasmid family that is distributed among *V. parahaemolyticus, V. harveyi,* and *V. campbellii* environmental strains. In addition, we show that plasmids are involved in the transfer of genes characteristic of *V. parahaemolyticus* clonal pandemic strains to environmental strains such as 22702. Phylogenetic analysis of *V. parahaemolyticus* 22702 in chapter 2 demonstrated the relatedness of this strain to *V. parahaemolyticus* clinical strains. Furthermore, in chapter 2 we demonstrate the existence of other emerging *V. parahaemolyticus* pathogens present among the environmental strains with the characterization of AF91, which is related to the clonal pandemic strains but lacks the virulence associated genes of the VPaI-7 genomic island. The characterization of p22702B in this chapter demonstrates the contribution of plasmids to horizontal transfer of genomic island genes associated with pandemic strains. The role of plasmids genomic islands in horizontal transfer of virulence-associated genes may explain the diversity of virulence profiles identified for disease-causing *V. parahaemolyticus* strains.

Acknowledgements

I would like to thank M. Barnstead, S. van Aken, G. Pai, and M. B. Craven for coordinating the library construction and sequencing of the plasmids. Also, S. Chen for initial research assistance characterizing the swarming phenotype. In addition, I would like to thank C. Lovell for providing some of the environmental strains examined in this study, B. Hammer for providing the pKAS46 suicide vector, and L. Williams for providing technical suggestions.

References

1. **(CDC), C. f. D. C. a. P.** 2007. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food--10 states, 2006. MMWR Morb. Mortal. Wkly. Rep. **55:**854-856.
2. **(CDC), C. f. D. C. a. P.** 2006. *Vibrio parahaemolyticus* infections associated with consumption of raw shellfish--three states, 2006. MMWR Morb. Mortal. Wkly. Rep. **55:**854-856.
3. **Agron, P. G., P. Sobecky, and G. L. Andersen.** 2002. Establishment of uncharacterized plasmids in *Escherichia coli* by in vitro transposition. FEMS Microbiol. Lett. **217:**249-54.
4. **Ahmed, A. M., S. Shinoda, and T. Shimamoto.** 2005. A variant type of *Vibrio cholerae* SXT element in a multidrug-resistant strain of *Vibrio fluvialis*. FEMS Microbiol. Lett. **242:**241-247.
5. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389-3402.
6. **Bagwell, C. E., Y. M. Piceno, A. Ashburne-Lucas, and C. R. Lovell.** 1998. Physiological diversity of the rhizosphere diazotroph assemblages of selected salt marsh grasses. Appl. Environ. Microbiol. **64:**4276-4282.
7. **Bairoch, A.** 1991. PROSITE: a dictionary of sites and patterns in proteins. Nucleic Acids Res. **19:**2241-2245.
8. **Balding, C., S. A. Bromley, R. W. Pickup, and J. R. Saunders.** 2005. Diversity of phage integrases in Enterobacteriaceae: development of markers for environmental analysis of temperate phages. Env. Microbiol. **7:**1558-1567.
9. **Bateman, A., L. Coin, R. Durbin, R. D. Finn, V. Hollich, S. Griffiths-Jones, A. Khanna, M. Marshall, S. Moxon, E. L. Sonnhammer, D. J. Studholme, C. Yeats, and S. R. Eddy.** 2004. The Pfam protein families database. Nucleic Acids Res. **32:**D138-D141.
10. **Bates, T. C., and J. D. Oliver.** 2004. The viable but nonculturable state of Kanagawa positive and negative strains of *Vibrio parahaemolyticus.* J. Microbiol. **42:**74-79.
11. **Besemer, J., and M. Borodovsky.** 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. **27:**3911-3920.
12. **Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453-74.
13. **Bose, M., and R. D. Barber.** 2006. Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. In Silico Biol. **6:**223-227.
14. **Boyd, E. F., A. L. V. Cohen, L. M. Naughton, D. W. Ussery, T. T. Binnewies, O. C. Stine, and M. A. Parent.** 2008. Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. BMC Microbiol. **8:**110-124.

15. **Breitbart, M., P. Salamon, B. Andresen, J. M. Mahaffy, A. M. Segall, D. Mead, F. Azam, and F. Rohwer.** 2002. Genomic analysis of uncultured marine viral communities. Proc. Natl. Acad. Sci. USA **99:**14250-14255.
16. **Brenner, D. J., F. W. Hickman-Brenner, J. V. Lee, A. G. Steigerwalt, G. R. Fanning, D. G. Hollis, J. J. Farmer 3rd, R. E. Weaver, S. W. Joseph, and R. J. Seidler.** 1983. *Vibrio furnissii* (formerly aerogenic biogroup of *Vibrio fluvialis*), a new species isolated from human feces and the environment. J. Clin. Microbiol. **18:**816-824.
17. **Buist, G., A. Steen, J. Kok, and O. P. Kuipers.** 2008. LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol. Microbiol. **68:**838-847.
18. **Carraturo, A., K. Raieta, D. Ottaviani, and G. L. Russo.** 2006. Inhibition of *Vibrio parahaemolyticus* by a bacteriocin-like inhibitory substance (BLIS) produced by *Vibrio mediterranei* I. J. Appl. Microbiol. **101:**234-241.
19. **Ceccarelli, D., A. M. Salvia, J. Sami, C. P., and M. M. Colombo.** 2006. New cluster of plasmid-located class 1 integrons in *Vibrio cholerae* O1 and a dfrA15 cassette-containing integron in *Vibrio parahaemolyticus* isolated in Angola. Antimicrob. Agents Chemother. **50:**2493-2499.
20. **Chang, B., H. Miyamoto, H. Taniguchi, and S. Yoshida.** 2002. Isolation and genetic characterization of a novel filamentous bacteriophage, a deleted form of phage f237, from a pandemic *Vibrio parahaemolyticus* O4:K68 strain. Microbiol. Immunol. **46:**565-569.
21. **Chang, B., H. Taniguchi, H. Miyamoto, and S. Yoshida.** 1998. Filamentous bacteriophages of *Vibrio parahaemolyticus* as a possible clue to genetic transmission. J. Bacteriol. **180:**5094-101.
22. **Chen, C. Y., Wu, K. M., Chang, Y. C., Chang, C. H., Tsai, H. C., Liao, T. H., Liu, Y. M., Chen, H. J., Shen, A. B., Li, J. C., Su, T. L., Shao, C. P., Lee, C. T., Hor, L. I., and Tsai, S. F.** 2003. Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. Genome Res. **13:**2577-87.
23. **Chowdhury, N. R., O. C. Stine, J. Glenn Morris, and G. B. Nair.** 2004. Assessment of evolution of pandemic *Vibrio parahaemolyticus* by multilocus sequence typing. J. Bacteriol. **42:**1280-1282.
24. **Clark, A. J., and A. D. Margulies.** 1965. Isolation and characterization of recombination-deficient mutants of *Escherichia coli* K12. Proc. Natl. Acad. Sci. USA **53:**451-459.
25. **Comeau, A. M., E. Buenaventura, and C. A. Suttle.** 2005. A persistent, productive, and seasonally dynamic vibriophage population within Pacific oysters (*Crassostrea gigas*). Appl. Environ. Microbiol. **71:**5324-5331.
26. **Comeau, A. M., A. M. Chan, and C. A. Suttle.** 2006. Genetic richness of vibriophages isolated in a coastal environment. Env. Microbiol. **8:**1164-76.
27. **Comeau, A. M., and C. A. Suttle.** 2007. Distribution, genetic richness, and phage sensitivity of *Vibrio* spp. from coastal British Columbia. Env. Microbiol. **9:**1790-1800.
28. **Cook, M. A., A. M. Osborn, J. Bettandorff, and P. A. Sobecky.** 2001. Endogenous isolation of replicon probes for assessing plasmid ecology of marine sediment microbial communities. Microbiology **147:**2089-2101.

29. **Criminger, J. D., T. H. Hazen, P. A. Sobecky, and C. R. Lovell.** 2007. Nitrogen fixation by *Vibrio parahaemolyticus* and its implications for a new ecological niche. Appl. Environ. Microbiol. **73:**5959-5961.
30. **Daniels, N. A., L. MacKinnon, R. Bishop, S. Altekruse, B. Ray, R. M. Hammond, S. Thompson, S. Wilson, N. H. Bean, P. M. Griffin, and L. Slutsker.** 2000. *Vibrio parahaemolyticus* infections in the United States, 1973-1998. J. Infect. Dis. **181:**1661-1666.
31. **DePaola, A., J. L. Nordstrom, A. Dalsgaard, A. Forslund, J. Oliver, T. Bates, K. L. Bourdage, and P. A. Gulig.** 2003. Analysis of *Vibrio vulnificus* from market oysters and septicemia cases for virulence markers. Appl. Environ. Microbiol. **69:**4006-4011.
32. **DePaola, A., J. Ulaszek, C. A. Kaysner, B. J. Tenge, J. L. Nordstrom, J. Wells, N. Puhr, and S. M. Gendel.** 2003. Molecular, serological, and virulence characteristics of *Vibrio parahaemolyticus* isolated from environmental, food, and clinical sources in North America and Asia. Appl. Environ. Microbiol. **69:**3999-4005.
33. **Di Lorenzo, M., M. Stork, M. E. Tolmasky, L. A. Actis, D. Farrell, T. J. Welch, L. M. Crosa, A. M. Wertheimer, Q. Chen, P. Salinas, L. Waldbeser, and J. H. Crosa.** 2003. Complete sequence of virulence plasmid pJM1 from the marine fish pathogen *Vibrio anguillarum* strain 775. J. Bacteriol. **185:**5822-5830.
34. **Dunn, A. K., M. O. Martin, and E. V. Stabb.** 2005. Characterizatioon of pES213, a small mobilizable plasmid from *Vibrio fischeri*. Plasmid **54:**114-134.
35. **Eisen, J. A.** 1995. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. J. Mol. Evol. **41:**1105-1123.
36. **Elliot, E. L., C. A. Kaysner, L. Jackson, and M. L. Tamplin.** 1998. *Vibrio cholerae, V. parahaemolyticus, V. vulnificus,* and other *Vibrio* spp., p. 9.01-9.27. *In* U. S. F. a. D. A. B. A. Manual (ed.). A.O.A.C. International, Gaithersburg, MD.
37. **Faruque, S. M., I. B. Naser, K. Fujihara, P. Diraphat, N. Chowdhury, M. Kamruzzaman, F. Qadri, S. Yamasaki, A. N. Ghosh, and J. J. Mekalanos.** 2005. Genomic sequence and receptor for the *Vibrio cholerae* phage KSF-1: evolutionary divergence among filamentous vibriophages mediating lateral gene transfer. J. Bacteriol. **187:**4095-4103.
38. **Garvey, P., A. Rince, C. Hill, and G. F. Fitzgerald.** 1997. Identification of a RecA homolog (RecALP) on the conjugative lactococcal phage resistance plasmid pNP40: evidence of a role for chromosomally encoded RecAL in abortive infection. Appl. Environ. Microbiol. **63:**1244-1251.
39. **González-Escalona, N., J. Martinez-Urtaza, J. Romero, R. T. Espejo, L. Jaykus, and A. DePaola.** 2008. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. J. Bacteriol. **190:**2831-2840.
40. **Gonzalez-Escalona, N., J. Romero, and R. T. Espejo.** 2005. Polymorphism and gene conversion of the 16S rRNA genes in the multiple rRNA operons of *Vibrio parahaemolyticus.* FEMS Microbiol. Lett. **246:**213-219.

41. **Gross, M., I. Marianovsky, and G. Glaser.** 2006. MazG--a regulator of programmed cell death in *Escherichia coli*. Mol. Microbiol. **59:**590-601.
42. **Hardies, S. C., A. M. Comeau, P. Serwer, and C. A. Suttle.** 2003. The complete sequence of marine bacteriophage VpV262 infecting *Vibrio parahaemolyticus* indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. Virology **310:**359-371.
43. **Hazen, T. H., K. D. Kennedy, S. Chen, S. V. Yi, and P. A. Sobecky.** 2009. Inactivation of mismatch repair increases the diversity of *Vibrio parahaemolyticus*. Env. Microbiol.**:**10.1111/j.1462-2920.2008.01853.x
44. **Hazen, T. H., T. M. Lowe, D. J. Silberger, P. C. Lafon, N. M. Garrett, M. M. Parsons, C. A. Bopp, and P. A. Sobecky.** In preparation. Molecular analysis of the emergence of pathogenic *Vibrio parahaemolyticus* strains from environmental populations.
45. **Hazen, T. H., D. Wu, J. A. Eisen, and P. A. Sobecky.** 2007. Sequence characterization and comparative analysis of three plasmids isolated from environmental *Vibrio* spp. Appl. Environ. Microbiol. **73:**7703-7710.
46. **Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter, and C. M. Fraser.** 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406:**477-83.
47. **Honma, Y., M. Ikema, C. Toma, M. Ehara, and M. Iwanaga.** 1997. Molecular analysis of a filamentous phage (fs1) of *Vibrio cholerae* O139. Biochim. Et Biophys. Acta **1362:**109-115.
48. **Hoyt, P. R., and R. K. Sizemore.** 1982. Competitive dominance by a bacteriocin-producing *Vibrio harveyi* strain. Appl. Environ. Microbiol. **44:**653-658.
49. **Hurley, C. C., A. M. Quirke, F. J. Reen, and E. F. Boyd.** 2006. Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates. BMC Genom. **7:**104.
50. **Ikema, M., and Y. Honma.** 1998. A novel filamentous phage, fs-2, of *Vibrio cholerae* O139. Microbiology **144:**1901-1906.
51. **Jensen, M. A., S. M. Faruque, J. J. Mekalanos, and B. R. Levin.** 2006. Modeling the role of bacteriophage in the control of cholera outbreaks. Proc. Natl. Acad. Sci. USA **103:**4652-4657.
52. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution of protein molecules., vol. Academic Press, New York.
53. **Kapfhammer, D., J. Blass, S. Evers, and J. Reidl.** 2002. *Vibrio cholerae* phage K139: complete genome sequence and comparative genomics of related phages. J. Bacteriol. **184:**6592-6601.
54. **Kieser, T.** 1984. Factors affecting the isolation of ccc DNA from *Streptomyces lividans* and *Escherichia coli*. Plasmid **12:**19-36.

55. **Kowalczykowski, S. C., D. A. Dixon, A. K. Eggleston, S. D. Lauder, and W. M. Rehrauer.** 1994. Biochemistry of homologous recombination in *Escherichia coli*. Microbiol. Rev. **58:**401-465.
56. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinf. **5:**150-163.
57. **Lee, J. V., P. Shread, A. L. Furniss, and T. N. Bryant.** 1981. Taxonomy and description of *Vibrio fluvialis* sp. nov. (synonym group F Vibrios, group EF6). J. Appl. Bacteriol. **50:**73-94.
58. **Limpens, E., C. Franken, P. Smit, J. Willemse, T. Bisseling, and R. Geurts.** 2003. LysM domain receptor kinases regulating rhizobial Nod factor-induced infection. Science **302:**630-633.
59. **Lindell, D., M. B. Sullivan, Z. I. Johnson, A. C. Tolonen, F. Rohwer, and S. Chisholm.** 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. Proc. Natl. Acad. Sci. USA **101:**11013-11018.
60. **Lobocka, M. B., D. J. Rose, G. Plunkett, M. Rusin, A. Samojedny, H. Lehnherr, M. B. Yarmolinsky, and F. R. Blattner.** 2004. Genome of bacteriophage P1. J. Bacteriol. **186:**7032-7068.
61. **Makino, K., K. Oshima, K. Kurokawa, K. Yokoyama, T. Uda, K. Tagomori, Y. Iijima, M. Najima, M. Nakano, A. Yamashita, Y. Kubota, S. Kimura, T. Yasunaga, T. Honda, H. Shinagawa, M. Hattori, and T. Iida.** 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. Lancet **361:**743-749.
62. **Martinez-Urtaza, J., A. Lozano-Leon, A. DePaola, M. Ishibashi, K. Shimada, M. Nishibuchi, and E. Liebana.** 2004. Characterization of pathogenic *Vibrio parahaemolyticus* isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates. J. Clin. Microbiol. **42:**4672-8.
63. **McCall, J. O., and R. K. Sizemore.** 1979. Description of a bacteriocinogenic plasmid in *Beneckea harveyi*. Appl. Environ. Microbiol. **38:**974-9.
64. **McDougald, D., S. Srinivasan, S. A. Rice, and S. Kjelleberg.** 2003. Signal-mediated cross-talk regulates stress adaptation in *Vibrio* species. Microbiology-Sgm **149:**1923-1933.
65. **Meador, C. E., M. A. Parsons, C. A. Bopp, P. Gerner-Smidt, J. A. Painter, and G. J. Vora.** 2007. Virulence gene- and pandemic group-specific marker profiling of clinical *Vibrio parahaemolyticus* isolates. J. Clin. Microbiol. **45:**1133-1139.
66. **Meibom, K. L., M. Blokesch, N. A. Dolganov, C. Y. Wu, and G. K. Schoolnik.** 2005. Chitin induces natural competence in *Vibrio cholerae*. Science **310:**1824-1827.
67. **Miller, E. S., J. F. Heidelberg, J. A. Eisen, W. C. Nelson, A. S. Durkin, A. Ciecko, T. V. Feldblyum, O. White, I. T. Paulsen, W. C. Nierman, J. Lee, B. Szczypinski, and C. M. Fraser.** 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: Comparative genomics of a T4-related bacteriophage. J. Bacteriol. **185:**5220-5233.
68. **Miller, M. C., D. P. Keymer, A. Avelar, A. B. Boehm, and G. K. Schoolnik.** 2007. Detection and transformation of genome segments that differ within a

coastal population of *Vibrio cholerae* strains. Appl. Environ. Microbiol. **73:**3695-7304.

69. **Murooka, Y., and T. Harada.** 1979. Expansion of the host range of coliphage P1 and gene transfer from enteric bacteria to other Gram-negative bacteria. Appl. Environ. Microbiol. **38:**754-757.
70. **Murphy, R. A., and E. F. Boyd.** 2007. Three pathogenicity islands of *Vibrio cholerae* can excise from the chromosome and form circular intermediates. J. Bacteriol.
71. **Nair, G. B., T. Ramamurthy, S. K. Bhattacharya, B. Dutta, Y. Takeda, and D. A. Sack.** 2007. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. Clin. Microbiol. Rev. **20:**39-48.
72. **Nasu, H., T. Iida, T. Sugahara, Y. Yamaichi, K. S. Park, K. Yokoyama, K. Makino, H. Shinagawa, and T. Honda.** 2000. A filamentous phage associated with recent pandemic *Vibrio parahaemolyticus* O3:K6 strains. J. Clin. Microbiol. **38:**2156-2161.
73. **Nei, M., and R. Chakraborty.** 1976. Empirical relationship between the number of nucleotide substitutions and interspecific identity of amino acid sequences in some proteins. J. Mol. Evol. **7:**313-323.
74. **Nishibuchi, M., and J. B. Kaper.** 1995. Thermostable direct hemolysin gene of *Vibrio parahaemolyticus*: a virulence gene acquired by a marine bacterium. Infect. Immun. **63:**2093-2099.
75. **Nystrom, T., K. Flardh, and S. Kjelleberg.** 1990. Responses to multiple-nutrient starvation in marine *Vibrio* sp strain-Ccug-15956. J. Bacteriol. **172:**7085-7097.
76. **Nystrom, T., R. M. Olsson, and S. Kjelleberg.** 1992. Survival, stress resistance, and alterations in protein expression in the marine *Vibrio* sp strain S14 during starvation for different individual nutrients. Appl. Environ. Microbiol. **58:**55-65.
77. **Oakey, H. J., B. R. Cullen, and L. Owens.** 2002. The complete nucleotide sequence of the *Vibrio harveyi* bacteriophage VHML. J. Appl. Microbiol. **93:**1089-1098.
78. **Okada, N., T. Iida, K. Park, N. Goto, T. Yasunaga, H. Hiyoshi, S. Matsuda, T. Kodama, and T. Honda.** 2009. Identification and characterization of a novel type III secretion system in *trh*-positive *Vibrio parahaemolyticus* strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level. Infect. Immun. **77:**904-913.
79. **Oliver, J. D., L. Nilsson, and S. Kjelleberg.** 1991. Formation of nonculturable *Vibrio vulnificus* cells and its relationship to the starvation state. Appl. Environ. Microbiol. **57:**2640-2644.
80. **Paludan-Müller, C., D. Weichart, D. McDougald, and S. Kjelleberg.** 1996. Analysis of starvation conditions that allow for prolonged culturability of *Vibrio vulnificus* at low temperature. Microbiology-Uk **142:**1675-1684.
81. **Park, K. S., T. Ono, M. Rokuda, M. H. Jang, K. Okada, T. Iida, and T. Honda.** 2004. Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. Infect. Immun. **72:**6659-6665.
82. **Parsons, M. B., K. L. Cooper, K. A. Kubota, N. Puhr, S. Simington, P. S. Calimlim, D. Schoonmaker-Bopp, C. Bopp, B. Swaminathan, P. Gerner-**

Smidt, and E. M. Ribot. 2007. PulseNet USA standardized pulsed-field gel electrophoresis protocol for subtyping of *Vibrio parahaemolyticus*. Foodborne Pathog. Dis. **4:**285-92.

83. **Pedersen, K., and J. L. Larsen.** 1995. Evidence for the existence of distinct populations of *Vibrio anguillarum* Serogroup O1 based on plasmid contents and ribotypes. Appl. Environ. Microbiol. **61:**2292-2296.
84. **Powers, L. G., J. T. Mallonee, and P. A. Sobecky.** 2000. Complete nucleotide sequence of a cryptic plasmid from the marine bacterium *Vibrio splendidus* and identification of open reading frames. Plasmid **43:**99-102.
85. **Prasad, S., P. C. Morris, R. Hansen, P. G. Meaden, and B. Austin.** 2005. A novel bacteriocin-like substance (BLIS) from a pathogenic strain of *Vibrio harveyi*. Microbiology **151:**3051-3058.
86. **Purdy, A., F. Rohwer, R. Edwards, F. Azam, and D. H. Bartlett.** 2005. A glimpse into the expanded genome content of *Vibrio cholerae* through identification of genes present in environmental strains. J. Bacteriol. **187:**2992-3001.
87. **Rashid, M. H., and A. Kornberg.** 2000. Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of *Pseudomonas aeruginosa*. Proc. Nat. Acad. of Sci. USA **97:**4885-4890.
88. **Reen, F. J., Almagro-Moreno, S., Ussery, D., and Boyd, E. F.** 2006. The genomic code: inferring *Vibrionaceae* niche specialization. Nat. Rev. Microbiol. **9:**697-704.
89. **Rubin, E. J., W. Lin, J. J. Mekalanos, and M. K. Waldor.** 1998. Replication and integration of a *Vibrio cholerae* cryptic plasmid linked to the CTX prophage. Mol. Microbiol. **28:**1247-1254.
90. **Ruby, E. G., M. Urbanowski, J. Campbell, A. Dunn, M. Faini, R. Gunsalus, P. Lostroh, C. Lupp, J. McCann, D. Millikan, A. Schaefer, E. Stabb, A. Stevens, K. Visick, C. Whistler, and E. P. Greenberg.** 2005. Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners. Proc. Natl. Acad. Sci. USA **102:**3004-3009.
91. **Schäffer, A. A., L. Aravind, T. L. Madden, S. Shavirin, J. L. Spouge, Y. I. Wolf, E. V. Koonin, and S. A. Altschul.** 2001. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. **29:**2994-3005.
92. **Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting.** 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. USA **95:**5857-5864.
93. **Seguritan, V., I. W. Feng, F. Rohwer, M. Swift, and A. M. Segall.** 2003. Genome sequences of two closely related *Vibrio parahaemolyticus* phages, VP16T and VP16C. J. Bacteriol. **185:**6434-6447.
94. **Shehane, S. D., and R. K. Sizemore.** 2002. Isolation and preliminary characterization of bacteriocins produced by *Vibrio vulnificus*. J. Appl. Microbiol. **92:**322-8.
95. **Skorupski, K., and R. K. Taylor.** 1996. Posistive selection vectors for allelic exchange. Gene **169:**47-52.

96. **Sobecky, P. A., T. J. Mincer, M. C. Chang, A. Toukdarian, and D. R. Helinski.** 1998. Isolation of broad-host-range replicons from marine sediment bacteria. Appl. Environ. Microbiol. **64:**2822-2830.
97. **Stine, O. C., S. Sozhamannan, Q. Gou, S. Zheng, J. G. J. Morris, and J. A. Johnson.** 2000. Phylogeny of *Vibrio cholerae* based on *recA* sequence. Infect. Immun. **68:**7180-7185.
98. **Story, R. M., I. T. Weber, and T. A. Steitz.** 1992. The structure of the *E. coli recA* protein monomer and polymer. Nature **355:**318-325.
99. **Stretton, S., S. J. Danon, S. Kjelleberg, and A. E. Goodman.** 1997. Changes in cell morphology and motility in the marine *Vibrio* sp Strain S14 during conditions of starvation and recovery. FEMS Microbiol. Lett. **146:**23-29.
100. **Tacket, C. O., F. Hickman, G. V. Pierce, and L. F. Mendoza.** 1982. Diarrhea associated with *Vibrio fluvialis* in the United States. J. Clin. Microbiol. **16:**991-992.
101. **Tarr, C. L., J. S. Patel, N. D. Puhr, E. G. Sowers, C. A. Bopp, and N. A. Strockbine.** 2007. Identification of *Vibrio* isolates by a multiplex PCR assay and *rpoB* sequence determination. J. Clin. Microbiol. **45:**134-40.
102. **Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale.** 2003. The COG database: an updated version includes eukaryotes. BMC Bioinf. **4:**41-55.
103. **Tatusov, R. L., E. V. Koonin, and D. J. Lipman.** 1997. A genomic perspective on protein families. Science **24:**631-637.
104. **Tatusova, T. A., and T. L. Madden.** 1999. Blast 2 sequences- a new tool for comparing protein and nucleotide sequences. FEMS Microbiol. Lett. **174:**247-250.
105. **Thompson, C. C., F. L. Thompson, K. Vandemeulebroecke, B. Hoste, P. Dawyndt, and J. Swings.** 2004. Use of *recA* as an alternative phylogenetic marker in the family *Vibrionaceae*. Int. J. Syst. Evol. Microbiol. **54:**919-924.
106. **Thompson, F. L., D. Gevers, C. C. Thompson, P. Dawyndt, S. Naser, B. Hoste, C. B. Munn, and J. Swings.** 2005. Phylogeny and molecular identification of *Vibrios* on the basis of multilocus sequence analysis. Appl. Environ. Microbiol. **71:**5107-5115.
107. **Thompson, F. L., T. Iida, and J. Swings.** 2004. Biodiversity of *Vibrios*. Microbiol. Mol. Biol. Rev. **68:**403-431.
108. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. Clustal-W- Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Res. **22:**4673-4680.
109. **Thompson, J. R., M. A. Randa, L. A. Marcelino, A. Tomita-Mitchell, E. Lim, and M. F. Polz.** 2004. Diversity and dynamics of a north Atlantic coastal *Vibrio* community. Appl. Environ. Microbiol. **70:**4103-4110.
110. **Turner, M. S., L. M. Hafner, T. Walsh, and P. M. Giffard.** 2004. Identification and characterization of the novel lysM domain-containing surface protein sep

from *Lactobacillus fermentum* BR11 and its use as a peptide fusion partner in *Lactobacillus* and *Lactococcus*. Appl. Environ. Microbiol. **70:**3673-3680.

111. **Vadivelu, J., S. D. Puthucheary, A. Mitin, C. Y. Wan, B. Van Melle, and J. A. Puthucheary.** 1996. Hemolysis and plasmid profiles of *Vibrio parahaemolyticus*. Southeast Asian J. of Trop. Med. Pub. Health **27:**126-131.
112. **Waldor, M. K., and J. J. Mekalanos.** 1996. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science **272:**1910-1914.
113. **Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay.** 1982. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases, and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO Journal **1:**945-951.
114. **Wommack, K. E., and R. R. Colwell.** 2000. Virioplankton: Viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. **64:**69-114.
115. **Wu, H., Y. Ma, Y. Zhang, and H. Zhang.** 2004. Complete sequence of virulence plasmid pEIB1 from the marine fish pathogen *Vibrio anguillarum* strain MVM425 and location of its replication region. J. Appl. Microbiol. **97:**1021-1028.

CHAPTER 4

INACTIVATION OF MISMATCH REPAIR INCREASES THE DIVERSITY OF *VIBRIO PARAHAEMOLYTICUS*

In this chapter, I present research describing the role of molecular mechanisms for increases in the accumulation of mutation and frequency of recombination that accelerates the evolution of *V. parahaemolyticus* under certain conditions. The objective of chapter 4 was to determine whether molecular mechanisms such as inactivation of DNA mismatch repair contributes to the diversity of *V. parahaemolyticus* strains. I hypothesized that inactivation of methyl-directed mismatch repair (MMR) occurs naturally among *V. parahaemolyticus* strains and increases the diversity of mutator strains, especially during environmental stress. The specific objectives are to: 1) to examine the occurrence of natural mutators among *V. parahaemolyticus* clinical and environmental strains, 2) confirm the role of inactivation of MMR for increased accumulation of spontaneous mutation observed for the natural mutator strains, and 3) determine whether mutator strains have increased genetic diversity.

Abstract

Inactivation of mismatch repair (MMR) has been shown to increase the accumulation of spontaneous mutations and frequency of recombination for diverse pathogenic bacteria. Currently, little is known regarding the role of mutator phenotypes for the diversification of natural populations of opportunistic human pathogens in marine environments. In this study, a higher frequency of mutators was detected among *V. parahaemolyticus* strains obtained from environmental sources compared to clinical sources. Inactivation of the MMR gene *mutS* caused increased antibiotic resistance and phase variation resulting in translucent colony morphologies. Increased nucleotide diversity in *mutS* and *rpoB* alleles from mutator compared to wild-type strains indicated a significant contribution of the mutator phenotype to the evolution of select genes. The results of this study indicate that the inactivation of MMR in *V. parahaemolyticus* leads to increased genetic and phenotypic diversity. This study is the first to report a higher frequency of natural mutators among *Vibrio* environmental strains and to provide evidence that inactivation of MMR increases the diversity of *V. parahaemolyticus*. In addition, sequence analysis of *opaR* from TR colonies compared to the wild-type OP colonies revealed an increased frequency of transition mutations indicating a potential contribution of inactivation of MMR to the nucleotide diversity of *opaR*. Furthermore, *V. parahaemolyticus* environmental strains exhibited increased phase variation during static growth conditions.

Introduction

Vibrio parahaemolyticus is a halophilic γ-proteobacterium that occurs in sediments, is associated with particles and zooplankton, and is found in high densities in mollusks residing in coastal environments (10, 56, 57). *V. parahaemolyticus* can cause gastroenteritis and less commonly wound infections in humans from either consumption of contaminated shellfish or infection of broken skin upon exposure to seawater. An increase in the frequency and the geographical range of *V. parahaemolyticus*-associated food-poisoning outbreaks has occurred in the United States in recent years (6) (37) and emphasizes the need to study the distribution and ecology of *V. parahaemolyticus* in coastal environments. *V. parahaemolyticus* isolates obtained from environmental samples have been shown to have considerable genetic and phenotypic diversity (16, 33, 34). Pulsed-field gel electrophoresis (PFGE) analysis of *V. parahaemolyticus* has demonstrated the clonal nature of clinical strains while environmental strains exhibit greater diversity based on DNA restriction pattern analysis (33, 45). In addition, a recent multilocus sequence typing (MLST) study that analyzed the nucleotide diversity of seven housekeeping genes (*recA, gyrB, pntA, dtdS, pyrC, dnaE,* and *tnaA*) further emphasized the clonality of the pandemic clinical strains while non-pandemic clinical strains and environmental strains had greater diversity in allelic profiles (16). Although *V. parahaemolyticus* strains were shown to have considerable genetic diversity, surprisingly little is known about the mechanisms and conditions that contribute to this observed diversity. Determining the mechanisms that increase the diversity of *V. parahaemolyticus* is important for understanding the conditions that contribute to the genetic diversification of environmental strains and the ability of these strains to acquire

novel phenotypes for survival and adaptation during changing environmental conditions or infection of a host.

Inactivation of methyl-directed mismatch repair (MMR) is one of several stress-induced mechanisms that increases genetic diversity (15, 26). The loss of function of MMR leads to the emergence of mutator strains characterized by high spontaneous mutation frequencies (9, 26). The MMR proteins, MutS and MutL, reduce the number of errors that accumulate during DNA replication by recognizing and removing mismatches, respectively (53). In addition, MMR limits the recombination of divergent sequences and a loss of function of MMR would lower barriers to recombination that would have otherwise restricted horizontal gene transfer (HGT) (35). For example, MMR-deficient strains of *Pseudomonas stutzeri* had increased recombination of divergent *rpoB* alleles (39). Inactivation of MMR was also shown to increase phenotypic diversity resulting in phase variation of *Neisseria meningitidis* (48, 49) and *P. aeruginosa* (30) thereby increasing the adaptive abilities of these strains. Mutators have been characterized for diverse pathogenic bacteria including *Escherichia coli* (29, 47), *P. aeruginosa* (43, 44), *Haemophilus influenzae* (58), *N. meningitidis* (48), *Salmonella* spp. (60), *Staphylococcus aureus* (46), and *Bacillus anthracis* (61). Many of these studies focused on the role of mutator strains for increased antibiotic resistance especially in regard to treating chronic infections. In contrast, few studies have examined the role of inactivation of MMR in increasing the diversity of bacteria in the environment.

The objective of this study was to determine whether *V. parahaemolyticus* strains of clinical or environmental origin exhibited mutator phenotypes. In addition, we determined whether inactivation of MMR increases the diversity of *V. parahaemolyticus*.

The nucleotide sequences of several conserved housekeeping genes from mutator and wild-type strains were analyzed to determine whether inactivation of MMR increases genetic diversity. This study provides the first characterization of natural mutators in environmental populations of *V. parahaemolyticus* and demonstrates a contribution of inactivation of MMR to the diversity of a marine microorganism.

Materials and Methods

Bacterial strains, media, and antibiotics. *V. parahaemolyticus* clinical strains used in this study were obtained from samples associated with human illnesses and were provided by the Centers for Disease Control and Prevention (CDC; Atlanta, GA) and characterized as previously described (38). *V. parahaemolyticus* environmental strains were isolated from sediment, water, and oysters collected in September 2006 from Skidaway Island, Georgia (n= 44) and Apalachicola Bay, Florida (n= 13) (Table 4.1). Additional environmental strains were isolated from North Inlet, NC (n= 21) as previously described (1) (Table 4.1). The *V. parahaemolyticus* environmental strains were isolated by plating sediment, water, or oyster samples on thiosulfate citrate bile salts sucrose agar (TCBS). The environmental strains from Georgia were isolated primarily from sediment and water associated with the plant *Spartina* spp. in the middle to low intertidal zones. Strains from Florida were isolated from sandy to muddy sediments and oysters collected from a seafood market. Environmental strains isolated from North Carolina were from sediments of the rhizosphere of *Juncus* spp. or *Spartina* spp. in a *Spartina*-dominated salt marsh (1).

Table 4.1. Bacterial strains and vectors used in this study

Strain/Vector	Serotype	Origin	Source
pKO2.0	N/A	N/A	Burns & DiChristina unpublished
pBBR1MCS	"	"	(36)
ß2155	"	"	(10)
EC100	"	"	"
V. parahaemolyticus RIMD 2210633	O3:K6	clinical	(45)
V. parahaemolyticus ATCC 17802	O3:K6	clinical	(18)
V. parahaemolyticus K1223	O3:K6	clinical	(50)
V. parahaemolyticus 22702	O5:Kuk	sediment	(7)
V. parahaemolyticus J-C1-5	unknown	sediment	(1)
V. parahaemolyticus J-C1-39	unknown	sediment	"
V. parahaemolyticus J-C2-27	unknown	sediment	"
V. parahaemolyticus J-C2-29	unknown	sediment	"
V. parahaemolyticus J-C2-34	unknown	sediment	"

Marine strains were grown at 30˚C in M10 broth (0.4% tryptone, 0.25% yeast extract, artificial seawater; 0.6 M NaCl, 0.02 M KCl, 0.1 M $MgSO_4$, 0.02M $CaCl_2$) or on M10 plates with 1.8% agar. *E. coli* strains were grown at 37˚C in Luria-Bertani (LB) broth with 1% NaCl or on LB plates with 1.5% agar. Phase variation was examined using heart infusion (HI) broth supplemented with 2% agar for plates (36). All strains used in this study are listed in Table 4.1. The antibiotics rifampin (Rif), ciprofloxacin (Cip), and chloramphenicol (Cm) were used at 100 µg/ml, 8 µg/ml, and 20 µg/ml; respectively, unless otherwise noted.

Spontaneous mutation. Spontaneous mutation frequencies to rif were determined as previously described (39, 43) with modifications for testing marine bacteria. Specifically, strains were grown overnight in M10 followed by serial dilutions and plating on M10 and M10rif. The mutation frequency was calculated as the CFU/ml on M10rif plates divided by CFU/ml on M10 without antibiotic. Plates were incubated

overnight at 30˚C. Each mutation frequency represents the average of five replicates. Mutator strains were characterized as those having at least a 10-fold elevated mutation frequency compared to the wild-type strain *V. parahaemolyticus* 17802. Mutation frequencies to cip were determined in the same manner as described for rif.

DNA amplification and sequencing. The 16S rRNA gene, *recA,* and *rpoA* nucleotide sequences for phylogenetic identification of the environmental strains were obtained as previously described (5, 20). All additional primers are listed in Table 4.2. We PCR-amplified and sequenced the predicted 2,613-bp coding region of *mutS* using primers 71F and 2710R (Table 4.2). Internal primers listed in Table 4.2 were used for sequencing. Primers to screen for insertions within the *mutS-rpoS* intergenic region are listed in Table 6. All PCR amplifications were performed using BioRad iproof HF polymerase with the GC buffer (Bio-Rad Laboratories; Hercules, CA) and standard reaction and cycle conditions. PCR amplicons were separated on a 0.7% agarose gel and DNA was extracted using the GenElute gel extraction kit (Sigma-Aldrich; St. Louis, MO). Sequencing was performed at the University of Nevada, Reno Genomics Facility and the Georgia Institute of Technology School of Biology Genome Center. Sequences were trimmed and contigs assembled using BioEdit (18).

Table 4.2. Primers for analysis of MMR

Primer	Sequence (5'- 3')	T_m	Expected amplicon (bp)	Source
mutS sequence analysis				
*mutS*71F	CGCTACGGAAATAAACAAAAGAACT	54	2,781	This study
*mutS*124F	AAACGAGCGTCTCAACTGCT	57	-	"
*mutS*169F	GCCGCTACGGAAATAAACAA	54	-	"
*mutS*492R	CATCATGGCTTCTTCGGTTT	53	-	"
*mutS*771F	TGATCGCCAAGATCACTCTG	55	-	"
*mutS*875R	AGAACTGCGGCGAGTGTATT	57	-	"
*mutS*1259F	ATGGTGGTGTGATTGCAGAA	55	1,451	"
*mutS*1478R	TTCTGCAATCACACCACCAT	54	1,507	"
*mutS*1629F	CCTAGCTTCTGCCGTTTCAC	56	-	"
*mutS*2440R	TGCTACATCCACAGTGCTTGG	58	-	"
*mutS*2710R	TGGATAACCATAGCCCTTTTCTGT	56	2,781	"
Intergenic region screen				
mutS-rpoS F	AAGCAAATCATACGGCTTGG	54	2,100	"
mutS-rpoS R	CCATTCGCTTGCCTATTCAT	53	2,100	"
Modification of pKO2.0				
pKO2.0IPCRAscI	GACT GGCGCGCC TTCGTTCAAGCCGAGATCGGCTTC	71		"
pKO2.0IPCRBspHI	GACT TCATGA AACATGAAACATCGACCCACGGCG	65		"
chlAscI	GACT GGCGCGCCAGCCGGAAGCATAAAGTGTAAAGC	70	1,195	"
chlBspHI	GACTTCATGATTTGGCGAAAATGAGACGTTGATC	61	1,195	"
Constructing *mutS* strain				
*mutS*D1ApaI	GACTGGGCCCTATCACCAAGGCGCATGATA	66	914	"
*mutS*D2fusion	CGATACAGCTCTTCAAGGGCCTGCATCATCGGAGTGTGTTT	68	914	"
*mutS*D3fusion	AAACACACTCCGATGATGCAGGCCCTTGAAGAGCTGTATCG	68	1,024	"
*mutS*D4SacI	GACTGAGCTCAAGAGTTTGACGCGAGCAGT	65	1,024	"
Confirming *mutS* deletion				
*mutS*D0	GTGCAGCAGCGATAAGCTC	56	4,200	"
*mutS*D5	GACAATGCGTCAGAGACGAC	56	4,200	"
Complementation				
*mutS*C1ApaI	GACTGGGCCCGTCCCGCAAGTAACACACCT	70	2,900	"
*mutS*C2XhoI	GACTCTCGAGGAAGTGGGACGAGAGATTGG	64	2,900	"
Phylogenetic markers				
*recA*33F	TGCGCTAGGTCAAATTGAAA	53	975	(5)
*recA*622R	TCGTTTCAGGGTTACCGAAC	55	-	"
*recA*1008R	AGCAGGTGCTTCTGGTTGAG	58	975	"
*rpoA*58F	AGCTCGACTCACGCAAAAGT	57	900	"
*rpoA*958R	CTAGGCGCATACCCAGAGAC	57	900	"
Sequence analysis of *rpoB*				
*rpoB*458F	AGGCGTGTTCTTCGACAGCGATAA	60	1,647	This study
*rpoB*1110F	GTAGAAATCTACCGCATGATG	50	963	(55)
*rpoB*2105R	CGGCTACGTTACGTTCGATACCAG	59	1,647	This study

*Amplicon size is indicated for primers used for PCR and not for internal sequencing primers

Sequence and phylogenetic analysis. Phylogenetic analysis was conducted to compare sequence variation of a concatenation of 16S rRNA gene, *recA,* and *rpoA* nucleotide sequences as previously described (5, 20). The nucleotide sequence lengths of 16S rRNA gene, *recA,* and *rpoA* used to construct the phylogenetic tree were 1,412-, 732-, and 723-bp, respectively. A neighbor-joining tree was constructed for the nucleotide sequences of the concatenated housekeeping genes in MEGA version 4.0 (28) using the Kimura-2 parameter model with 1,000 bootstrap replications. Alignments were performed in BioEdit (18) and MEGA (28) and all phylogenetic analyses were conducted in MEGA (28). The nucleotide diversity (π), nucleotide divergence (θ_W), and the K_a/K_s ratio were determined using DnaSP version 4.5 (50) relative to *V. harveyi* as an outgroup. Sawyer's test in MEGA (28) and DnaSP (50) were used to test for evidence of recombination.

Construction of a Δ*mutS* strain. Deletion of >90% of the predicted coding region of *mutS* was performed using methods previously developed (J. L. Burns and T. J. DiChristina; submitted for publication) with adaptations for use with *V. parahaemolyticus*. In order to increase the efficiency of in-frame deletions in *V. parahaemolyticus,* the suicide vector pKO2.0 was modified using inverse PCR to remove the gentamycin resistance cassette and attach the restriction sites AscI and BspHI to the remaining vector using the primers listed in Table 6. The chloramphenicol acetyl-transferase (CAT) from pBBR1MCS (27) was PCR-amplified and restriction sites AscI and BspHI were added for ligation into the inverse PCR-amplified pKO2.0. The CAT amplicon and pKO2.0 were digested with AscI and BspHI then gel-purified and ligated together using the Fast-Link DNA ligation kit (Epicentre; Madison, WI). The assembled vector was electroporated into *E. coli*

EC100 and isolated using the Qiagen mini-prep kit (Qiagen; Valencia, CA) followed by restriction endonuclease digestion with AscI and BspHI to confirm the ligation.

The *mutS* deletion was performed in *V. parahaemolyticus* ATCC 17802, which is an O3:K6 strain (14). Sequence analysis of *mutS* and adjacent DNA in 17802 showed this region was identical to that of the completed genome of *V. parahaemolyticus* RIMD2210633 (31). The 17802 and RIMD sequence data were used to design all primers. Following modification of pKO2.0 we used primers to amplify 894- and 1,003-bp flanking regions adjacent to *mutS* in 17802 and attach restriction sites SacI and ApaI. These amplicons were fused in a PCR reaction using the external primers D1 and D4. Fusion products were digested with ApaI and SacI then ligated into pKO2.0 as described above. The ligation was then transformed into *E. coli* EC100 (Table 4.2) and confirmed by restriction endonuclease digestion. The assembled suicide vector pKO2.0mutS was transformed into *E. coli* β2155 (7) (Table 4.2) and plated on LB supplemented with 20 μg/ml chloramphenicol and 100 μg/ml diaminopimelic acid (dap). pKO2.0mutS was transferred from β2155 to17802 by bi-parental mating. Colonies that were resistant to chloramphenicol and grew without dap were screened for a single recombination event using primers D0 and D4 and primers D1 and D5 (Table 4.2). A confirmed single-cross was grown overnight in LB with 0.25% NaCl and plated on 10% sucrose 0.25% NaCl LB plates. *V. parahaemolyticus* as well as other *Vibrios* are unable to grow without NaCl, and including NaCl in the sucrose plates reduces the efficiency of using *sacB* as counterselection. As previously reported for *V. anguillarum,* the *V. parahaemolyticus* colonies containing the vector initially grew then exhibited delayed sucrose sensitivity (40). Strains containing the suicide vector were plated on sucrose and incubated for 36-

48 h. Colonies were then screened for wild-type morphology as an indication the suicide vector had been lost as previously described (40). Colonies that exhibited a wild-type morphology and were sensitive to chloramphenicol were examined for deletion of *mutS* using primers D0 and D5 (Table 4.2). This PCR generated an approximately 1.9-kb amplicon compared to the 4.2-kb wild-type amplicon. The amplicon from the mutant was sequenced to confirm that the deletion was in-frame. The deletion of *mutS* was also verified by an inability to PCR-amplify using the internal primers 1259F and 2440R (Table 4.2).

Complementation with wild-type *mutS.* The Δ*mutS* strain was confirmed to have a *mutS* deletion by complementation with wt *mutS* PCR amplified from 17802 with primers *mutS*C1 and *mutS*C2 (Table 4.2). The gel-purified product was restriction digested with XhoI and ApaI and ligated into XhoI and ApaI digested and purified pBBR1MCS (Table 4.2). The ligation was transformed into *E. coli* EC100. The DNA was recovered and transformed into *E. coli* β2155 then transferred to each of the mutator and wild-type strains by bi-parental mating as described above. Five chloramphenicol resistant colonies for each strain were assayed for changes in spontaneous mutation to rif as described above. J-C2-34 was also complemented with *mutH* and *mutL* PCR-amplified using primers *mutH*C1 and *mutH*C2 and primers *mutL*C1 and *mutL*C2 (Table 4.2).

Analysis of *rpoB* from Rif^r colonies. Sequence analysis of *rpoB* from Rif^r colonies was performed by PCR-amplification of partial *rpoB* regions from colonies that had grown overnight on Rif. The DNA used for PCR-amplification was isolated directly from Rif^r colonies and amplified using conditions described above with primers *rpoB*458F and *rpoB*2105R from a previous study (55) listed in Table 4.2. A total of 10 colonies were

examined for each of the strains 17802, Δ*mutS*, J-C1-5, and J-C2-29. Sequences were aligned as described above and the types of mutations were determined by visual inspection.

Percentage of translucent colonies. The frequency of phase variation was examined by plating strains on heart infusion (HI) agar (Difco), which was previously used to distinguish between opaque (OP) and translucent (TR) colony phenotypes (36) to monitor opacity. Strains were grown overnight in HI broth at 30˚C with shaking and plated on HI agar. The percentage of TR CFU/ml was determined by dividing the TR CFU/ml by the combined total of TR and OP CFU/ml on HI agar. Values represent the averages of 5-7 independent replicates for each strain.

Results

V. parahaemolyticus MMR homologs and organization of the *mutS-rpoS* region. To determine if homologs to known MMR proteins were present in *V. parahaemolyticus*, we examined the sequenced genome of the pandemic O3:K6 *V. parahaemolyticus* RIMD2210633. A BLAST search of the genome of RIMD2210633 using MMR proteins of *E. coli* revealed *V. parahaemolyticus* possessed homologs of all the major MMR proteins (Table 4.3). *V. parahaemolyticus* RIMD2210633 did not have a homolog of the *V. cholerae* MMR protein MutK, which to date has only been identified in *V. cholerae* (2). In *E. coli* K12 and other Gram-negative bacteria, *mutS* is located adjacent to the stationary phase regulator *rpoS* (12) (Figure 4.1). The *mutS* of Gram-positive bacteria such as *S. aureus* is organized as a *mutSL* operon (46) rather than the *mutS-rpoS* organization characterized for most Gram-negative bacteria. In addition, the *mutS-rpoS*

intergenic region of *E. coli* was previously characterized as a site for recombinational insertion of DNA segments of 8.5 to 14.5-kb (21) (Figure 4.1). Although the *V. parahaemolyticus* RIMD2210633 genome had a similar genetic organization of *mutS* to that of other Gram-negative bacteria, this strain lacked insertions in the intergenic region as previously reported for *E. coli* (21) (data not shown). To determine whether there were insertions within the *mutS-rpoS* intergenic region among our collection of clinical and environmental strains, we designed primers to amplify the *mutS-rpoS* intergenic region. A PCR screen of the nine clinical and environmental strains using primers designed in this study indicated there were no detectable insertions within this intergenic region (data not shown).

Table 4.3. Mismatch repair homologs of *V. parahaemolyticus* RIMD2210633

Gene	Gene Id	Coordinates (Chr)
mutS	VP2552	2695120-2697681 (I)
mutK	-	-
mutL	VP2819	2984301-2982292 (I)
mutH	VP0518	535688-535008 (I)
mutU/uvrD	VP0041	47155-49170 (I)
	VP1083	1135576-1138995 (I)
	VP3013	3216340-3218514(I)
mutT	VP0468	473012-473410 (I)
	VPA0758	789046-789504 (II)

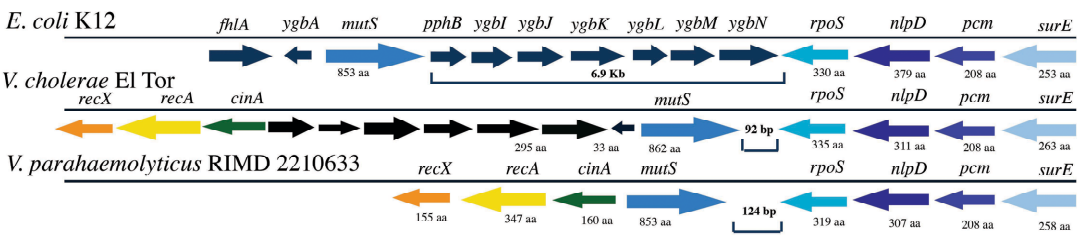


Figure 4.1. Genetic organization of the *mutS-rpoS* region of *E. coli*, *V. cholerae*, and *V. parahaemolyticus.* The size (bp) of the *mutS-rpoS* intergenic region is indicated for each strain.

Identification of natural mutators among *V. parahaemolyticus* environmental strains. *V. parahaemolyticus* strains from clinical and environmental sources (n= 142) were examined for possible mutator phenotypes that may have arisen due to inactivation of MMR. Of the 78 environmental and 64 clinical *V. parahaemolyticus* strains tested, four environmental *V. parahaemolyticus* strains exhibited at least 10-fold greater mutation frequency to Rif compared to the wild-type strain *V. parahaemolyticus* ATCC 17802 (Table 4.4). The majority of the *V. parahaemolyticus* clinical strains were isolated from different outbreaks that occurred in the U.S. from 1997-2007 (Table 4.1). Among the clinical strains are 20 with the clonal pandemic serotypes O3:K6 and O4:K12 while the remaining clinical strains represent diverse serotypes. Surprisingly, mutator phenotypes were not identified among any of the *V. parahaemolyticus* clinical strains tested. The environmental strains were isolated from sediment and water associated with the saltmarsh plants *Juncus* spp. of the high intertidal zone and *Spartina* spp. from the middle to low intertidal zone. All of the *V. parahaemolyticus* mutator strains were isolated from

the rhizosphere of the saltmarsh plant *Juncus* spp. from North Inlet, North Carolina. There was an additional mutator (strain SG358; Table 4.4) most closely related to *V. campbellii* that was isolated from the water column of a *Spartina*-dominated saltmarsh in Georgia. Sequence analysis of partial 16S rDNA, *rpoA,* and *recA* alleles from this strain revealed to 100, 99, and 98% nucleotide identity to *V. campbellii,* respectively.

Table 4.4. Frequency of spontaneous mutations conferring antibiotic resistance among wild-type and mutator strains

	Rifampin				Ciprofloxacin		
Strain[a]	Avg ± SD[b]	Fold difference[c]	Complemented with wt *mutS*	Fold change[d]	Avg ± SD[b]	Fold difference[c]	Complemented with wt *mutS*
17802	$7.7 \pm 6.5 \times 10^{-9}$	-	$6.7 \pm 4.1 \times 10^{-9}$	-0.13	$9.1 \pm 2.8 \times 10^{-7}$	-	-
17802 *ΔmutS*	**$8.4 \pm 2.1 \times 10^{-7}$**	**108.1**	**$7.6 \pm 6.4 \times 10^{-8}$**	**-10.05**	**$9.3 \pm 3.4 \times 10^{-5}$**	**101.2**	**$5.0 \pm 4.1 \times 10^{-6}$**
RIMD	$1.0 \pm 0.8 \times 10^{-8}$	0.3	$7.7 \pm 5.4 \times 10^{-9}$	-0.30	$2.1 \pm 0.9 \times 10^{-7}$	-0.8	ND
K1223	$1.0 \pm 0.6 \times 10^{-8}$	0.3	$6.0 \pm 4.1 \times 10^{-9}$	-0.67	$1.2 \pm 0.6 \times 10^{-6}$	0.3	ND
22702	$1.0 \pm 0.9 \times 10^{-8}$	0.3	$9.7 \pm 9.9 \times 10^{-9}$	-0.03	$1.5 \pm 0.4 \times 10^{-7}$	-0.8	ND
J-C1-5	**$6.6 \pm 1.7 \times 10^{-7}$**	**84.7**	**$9.4 \pm 3.8 \times 10^{-9}$**	**-69.21**	**$3.0 \pm 3.0 \times 10^{-5}$**	**32.0**	ND
J-C1-39	**$1.3 \pm 0.9 \times 10^{-6}$**	**167.8**	**$6.3 \pm 3.9 \times 10^{-9}$**	**-205.35**	**$3.2 \pm 3.4 \times 10^{-7}$**	**-0.6**	ND
J-C2-27	**$6.2 \pm 4.8 \times 10^{-7}$**	**79.5**	**$3.8 \pm 2.7 \times 10^{-9}$**	**-162.16**	**$4.1 \pm 3.9 \times 10^{-7}$**	**-0.5**	ND
J-C2-29	$5.0 \pm 3.0 \times 10^{-9}$	-0.4	$12.4 \pm 6.2 \times 10^{-9}$	-1.48	$3.0 \pm 1.3 \times 10^{-7}$	-0.7	ND
J-C2-34	**$1.5 \pm 0.8 \times 10^{-6}$**	**193.8**	**$2.5 \pm 1.1 \times 10^{-6}$**	**0.67**	**$6.9 \pm 5.7 \times 10^{-7}$**	**-0.2**	ND
SG358	**$6.3 \pm 4.2 \times 10^{-7}$**	**80.8**	**$4.9 \pm 2.8 \times 10^{-7}$**	**-0.29**	**$4.3 \pm 0.9 \times 10^{-5}$**	**46.3**	ND

In addition to determining whether *V. parahaemolyticus* strains had increased mutations conferring resistance to Rif, we determined whether they also had increased resistance to Cip (Table 4.4). One of the *V. parahaemolyticus* mutator strains (J-C1-5) and the *V. campbellii* mutator strain SG358 had higher resistance to Cip (3.0×10^{-5} and 4.3×10^{-5} respectively; Table 4.4) relative to 17802.

Deletion of *mutS* increased resistance while complementation of natural mutator strains with wild-type *mutS* restored wild-type sensitivities. The role of inactivation of *mutS* for increased accumulation of mutation conferring resistance to Rif and Cip was determined by construction of a 17802 Δ*mutS* strain. Deletion of *mutS* in 17802 resulted in a 108-fold higher mutation frequency compared to the wild-type strain conferring Rif^r (Table 4.4). Likewise, the Δ*mutS* strain had a 101-fold higher mutation frequency conferring resistance to Cip (Table 4.4). To determine if inactivation of *mutS* was a possible cause for the spontaneous mutation frequencies detected in the mutator strains each mutator was complemented with the wild-type *mutS* from 17802. The *ΔmutS* strain and three of the mutator strains were restored to wild-type mutation frequencies following complementation with wild-type *mutS* (Table 4.4). In contrast, the mutator strain J-C2-34 exhibited little change in Rif^r when complemented with the wild-type *mutS* (Table 4.4). In order to determine whether inactivation of other MMR genes may be the cause of the Rif^r in J-C2-34 we complemented this strain with *mutH* and *mutL* from 17802. Complementation of J-C2-34 with *mutH* and *mutL* resulted in no change in the spontaneous mutation frequency to Rif^r (data not shown). This result suggests that other MMR genes such as *uvrD* may have been inactivated or multiple MMR genes may be simultaneously defective. In addition, the *V. campbellii* mutator strain SG358, was not

complemented with the *mutS* from *V. parahaemolyticus* 17802 (Table 4.4). The MMR gene *mutH* of *V. cholerae* was previously shown to restore wild-type function to an *E. coli mutH* mutant indicating the conserved function of these proteins (13). The lack of complementation of *V. campbellii* SG358 in this study indicates that *mutS* of *V. parahaemolyticus* may not be transferable to closely related *Vibrios* or that other MMR genes may be defective in SG358.

Increased diversity of *mutS* and *rpoB* alleles from mutator strains. A neighbor-joining tree was constructed from a concatenation of the housekeeping gene nucleotide sequences to examine the diversity of the *V. parahaemolyticus* wild-type and mutator strains characterized in this study. The *V. parahaemolyticus* strains formed a monophyletic group that consisted of separate clusters of environmental and clinical strains (Figure 4.2) with the exception of the environmental strain 22702. *V. parahaemolyticus* 22702 was previously isolated from a saltmarsh environment and is able to fix nitrogen (5) yet clustered with the strains of clinical origin (Figure 4.2). The two *V. parahaemolyticus* clusters were distinct from the closely related *V. alginolyticus, V. harveyi,* and *V. campbellii* sequences indicating that they formed separate clusters within the *V. parahaemolyticus* group (Figure 4.2). Likewise, a phylogenetic tree of the *mutS* nucleotide sequences had a monophyletic organization for *V. parahaemolyticus* similar to that observed with the concatenated tree (Figure 4.3). In addition, the *mutS* sequences from three of the mutator strains formed a distinct group (Figure 4.3).

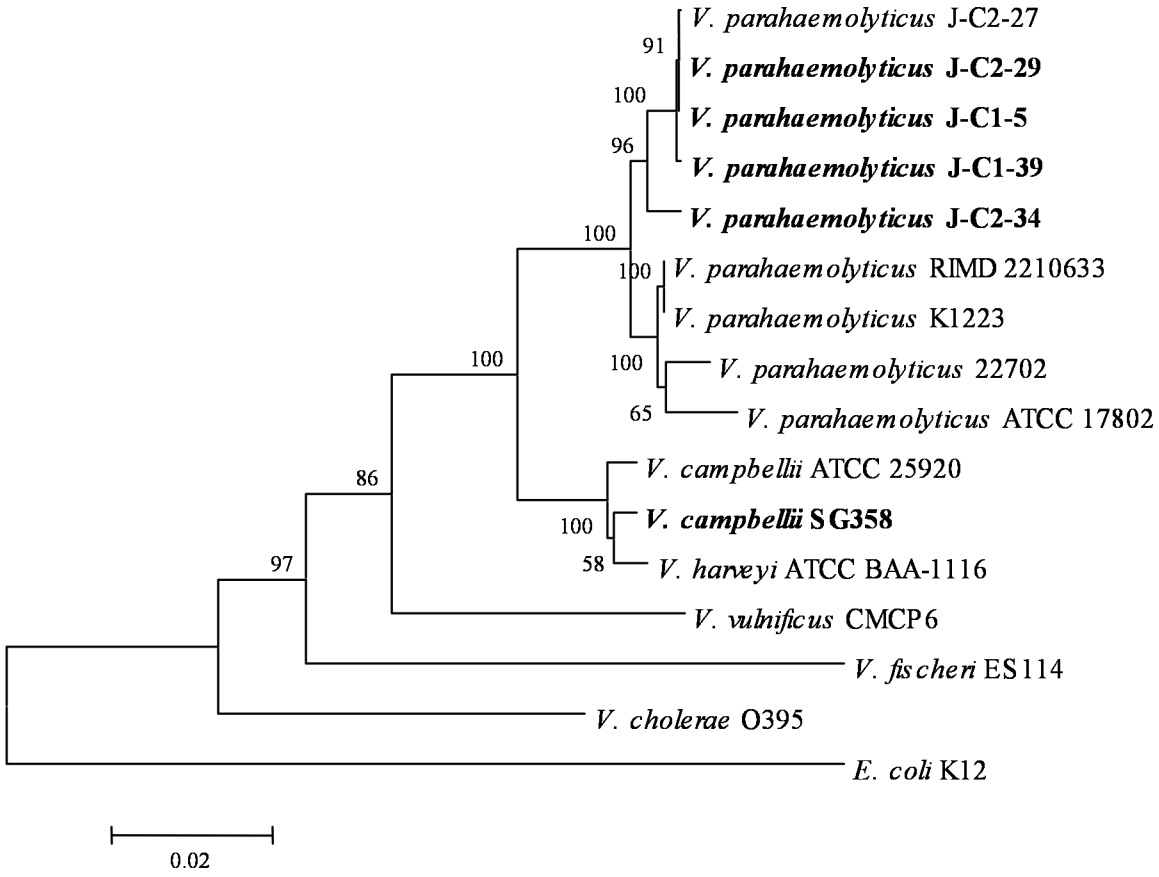


Figure 4.2. Neighbor-joining tree of a concatenation of 16S rRNA gene (1,412-bp), *recA* (732-bp), and *rpoA* (723-bp) nucleotide sequences from *V. parahaemolyticus* clinical and environmental strains compared to other *Vibrios*. The tree was constructed with the Kimura 2-parameter model and values shown represent 1,000 bootstrap replications. The scale bar represents 0.02 nucleotide substitutions per site. Mutator strains are indicated in bold.

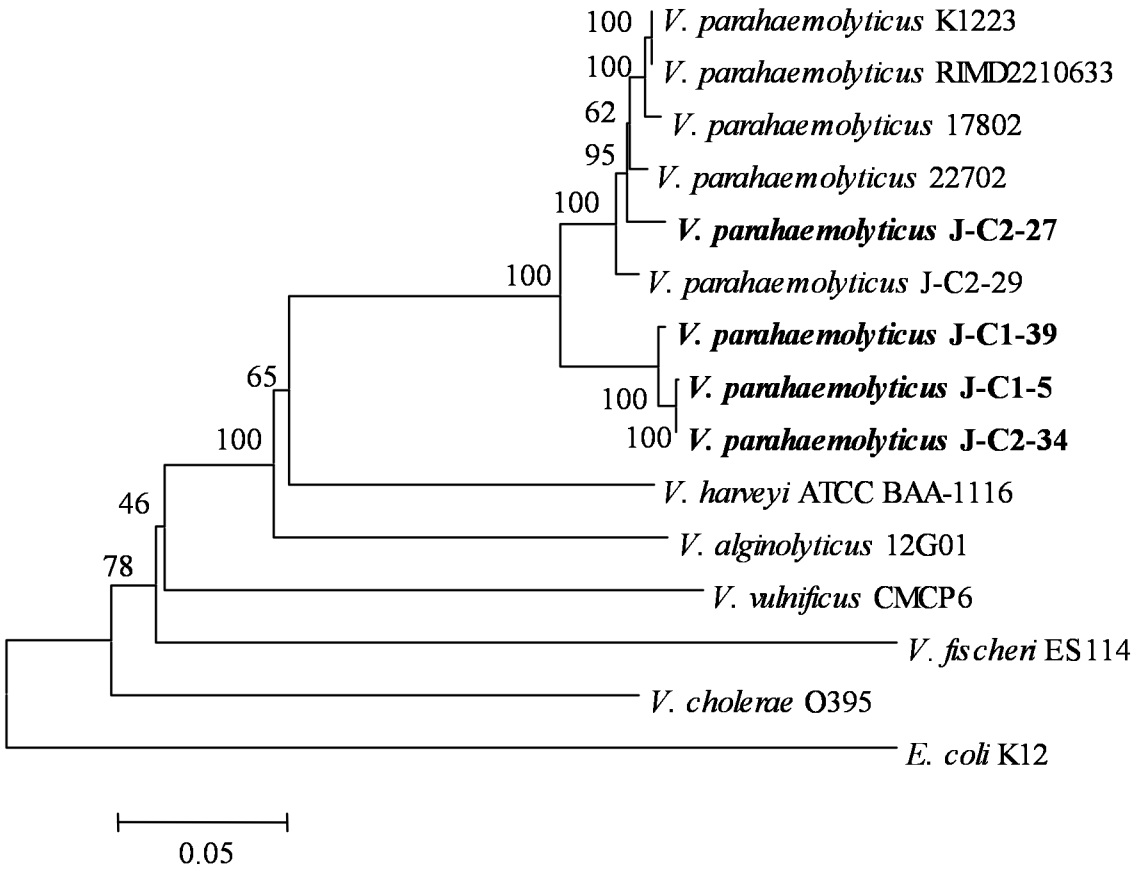


Figure 4.3. Neighbor-joining tree of the *mutS* nucleotide sequences (2,490-bp) from *V. parahaemolyticus* clinical and environmental strains compared to sequences available in GenBank for other *Vibrios*. The tree was constructed with the Kimura 2-parameter model and values represent 1,000 bootstrap replications. The scale bar represents 0.05 nucleotide substitutions per site. Mutator strains are indicated in bold.

To compare the level of genetic diversity of wild-type and mutator strains, we analyzed sequence diversity of the following four loci: *recA, rpoA, rpoB,* and *mutS*. We sequenced five and four alleles from the wild-type and mutator strains, respectively. The lengths of the nucleotide sequences analyzed were 729-, 723-, 1,539-, and 2,487-bp for *recA, rpoA, rpoB,* and *mutS,* respectively (Table 4.5). Also, the MutS coding region was divided into smaller regions that encompassed the MutS I (pfam01624) and MutS II (pfam05188) protein domains (nt 22-759), MutS III protein domain (nt 787-1,689), and the MutS ATPase domain (1,732-2,376) (Table 4.5). The nucleotide diversity within each domain was assessed independently to determine whether there are regions within the *mutS* gene that are undergoing increased mutation and/or recombination. We analyzed the level of diversity at the nucleotide level using two commonly used measures, the nucleotide diversity (π), (41) and Watterson's theta (θ_W), (59). The level of nucleotide diversity for nonsynonymous and synonymous sites was examined separately to determine whether different neutral and selective forces are influencing these two types of sites (Table 4.5).

Table 4.5. Nucleotide diversity of housekeeping genes from *V. parahaemolyticus* wild-type and mutator strains

Gene	Length (bp)	No. polymorphic sites	θ_W Total[a]	π[a]			Divergence from *V. harveyi*[a]	Z-test p-value[a]	Ka/Ks[a]	SSCF (p-value)[b]	SSUF (p-value)[b]	No. of recombination events[a]
				Synonymous	Non-synonymous	Total						
recA	729	40	0.02116 (0.01352)	0.07796 (0.04857)	0.00073 (0.00091)	0.01928 (0.01240)	0.10100 (0.10202)	0.00 (0.00)	0.004 (0.006)	1,879 (0.0006)	1,027,829 (0.6481)	2
rpoA	723	10	0.00465 (0.00302)	0.01454 (0.00839)	0.00074 (0.00092)	0.00415 (0.00277)	0.03714 (0.03758)	0.02 (0.12)	0.014 (0.015)	276 (0.5058)	3,436,349 (0.4983)	0
rpoB	1,539	53	0.00998 (0.01240)	0.03571 (0.03940)	0.00069 (0.00257)	0.00916 (0.01148)	0.06060 (0.06247)	0.00 (0.00)	0.071 (0.066)	5,543 (0.0026)	9,779,819 (0.0232)	5
mutS	2,487	198	0.01255 (0.03816)	0.04468 (0.14112)	0.00132 (0.00291)	0.01174 (0.03612)	0.21185 (0.21519)	0.00 (0.00)	0.023 (0.024)	95,500 (0.0000)	20,156,770 (0.0000)	8
mutS I & II	720	44	0.01333 (0.02879)	0.04387 (0.10227)	0.00147 (0.00184)	0.01181 (0.02639)	0.20364 (0.19521)	0.00 (0.00)	0.012 (0.014)	3,306 (0.0000)	1,059,640 (0.0556)	3
mutS III	903	54	0.01435 (0.02537)	0.05485 (0.09878)	0.00058 (0.00363)	0.01351 (0.02621)	0.23383 (0.23929)	0.00 (0.00)	0.028 (0.031)	7,762 (0.0000)	3,355,901 (0.0003)	2
mutS ATPase	645	74	0.00670 (0.05751)	0.02445 (0.20889)	0.00082 (0.00306)	0.00651 (0.05271)	0.17760 (0.19399)	0.01 (0.00)	0.018 (0.017)	8,506 (0.5871)	750,742 (0.9949)	0

[a]Values are reported independently for wild-type and (mutator) sequences

The nucleotide diversities of *rpoA* and *recA* loci were similar between the mutator compared to the wild-type strains (Table 4.5). In contrast, the nucleotide diversity was greatly elevated in mutator strains compared to wild-type strains for *mutS* and *rpoB*. This was evident for both synonymous and nonsynonymous sites (Table 4.5), suggesting that there is an overall increase of mutation rates in *mutS* and *rpoB* of mutator strains. The mutS ATPase domain showed the largest increase of nucleotide diversity in mutator compared to wild-type strains (Table 4.5). Synonymous sites of the mutator strains were approximately 10-fold more variable than those of wild-type strains in this locus. Visual inspection of the *mutS* alleles did not indicate any frameshift or nonsense mutations; however, there were numerous non-synonymous changes (data not shown).

Despite the general increase of the nucleotide variability in mutator compared to wild-type strains, mutator alleles appear to be under strong selective constraint. One method to examine the presence of selective constraint in protein-coding sequences is by comparing the ratio of nonsynonymous to synonymous site changes. Here we examined the ratio of nonsynonymous to synonymous changes in the four loci at two levels. First, we examined the selective constraint acting on the population by analyzing levels of nucleotide diversity (π), which measures within population diversity. Second, we examined the ratio of nonsynonymous to synonymous substitutions to a *V. harveyi* allele (K_a/K_s). The K_a/K_s relative to *V. harveyi* is indicative of selective constraint on an evolutionary timescale since the divergence of *V. harveyi* and *V. parahaemolyticus.* The ratios of nucleotide diversity for nonsynonymous to synonymous sites (π_{nonsyn}/π_{syn}) for mutator alleles were all below 1 and comparable to those in wild-type alleles, indicating the presence of strong purifying selection at the population level. For example, even

though overall nucleotide diversity for the *mutS* ATPase is almost 10-fold greater in mutator strains than in wild-type strains, π_{nonsyn}/π_{syn} in the mutator strains is even lower than that in the wild-type strains (0.0146 versus 0.0335). This pattern is concordant with the long-term selective constraint measured by the ratio of nonsynonymous to synonymous substitutions (K_a/K_s). For all four of the genes analyzed, the K_a/K_s (compared to a *V. harveyi* allele) ratios are all below 1, indicating the presence of strong purifying selection (Table 4.5). The Z-test for neutral selection relative to *V. harveyi* also confirmed that both the mutator and wild-type alleles were under purifying selection for each gene ($p<0.05$; Table 4.5).

Recombination was detected using two tests. First, we used the Sawyer's test, implemented in the DnaSP program (50). This test examines whether there are more consecutive identical polymorphic sites than expected, by calculating the sum of squared lengths of condensed fragments (SSCF) and the sum of squared lengths of uncondensed fragments (SSUF) (51). The significance of the SSCF and SSUF scores were obtained by means of 10,000 random simulations. It is known that Sawyer's test may be too stringent in determining recombination events (52). Thus, we used an additional test, the four-gametic test of Hudson and Kaplan (23) to detect recombination. Both tests indicated there was evidence of recombination in *recA, rpoB,* and *mutS* (Table 4.5). The MutS I and II, and MutS III domains had undergone recombination while the MutS ATPase domain had not (Table 4.5). When we do not separate the MutS locus into three domains, the number of total recombination events in this locus increases to eight. This suggests that we may have lost some power to detect recombination events between domains by separating the MutS coding region into three domains (Table 4.5).

Increased transition mutations in *rpoB* of Rif^r mutator strains. Partial sequences of *rpoB* from Rif^r colonies were examined from the wild-type strain 17802, the 17802 Δ*mutS* strain, mutator strain J-C1-5, and wild-type strain J-C2-29 to determine whether inactivation of MMR increased the frequency of transition mutations among mutator strains. Sequence analysis of a 995-bp region of *rpoB* from 10 Rif^r colonies each of 17802, Δ*mutS,* J-C1-5, and J-C2-29 revealed six potential sites that are hotspots of mutation for *V. parahaemolyticus* (Table 4.6). The nucleotide locations are indicated relative to the start of the *V. parahaemolyticus* coding region (Table 4.6). All of the mutation sites resulted in an amino acid change (Table 4.6). The percentage of nucleotide changes that were transitions represented 100% of the overall nucleotide differences for the mutator strains compared to only 10% and 40% for the wild-type strains (Table 4.6).

Table 4.6. Mutations in *rpoB* of Rif^r colonies from *V. parahaemolyticus* wild-type and mutator strains

Mutation type	Nucleotide*	Amino acid change	17802	***ΔmutS***	**J-C1-5**	J-C2-29
G-A	1586	R529H	1	**7**	**0**	0
A-G	1538	Q513R	0	**0**	**2**	0
	1577	H526R	0	**0**	**6**	1
	1703	N586S	0	**0**	**2**	0
C-T	1576	H526Y	0	**0**	**2**	2
	1592	S531F	0	**3**	**0**	1
Total Transitions			1	**10**	**12**	4
A-T	1538	Q513L	8	**0**	**0**	5
A-C	1538	Q513P	0	**0**	**0**	1
C-A	1592	S531Y	1	**0**	**0**	0
Total Transversions			9	**0**	**0**	6
Total Mutations			10	**10**	**12**	10
% Transitions			10	**100**	**100**	40

*All nucleotide locations are designated according to the RIMD2210633 *rpoB* coding region.

Phenotypic variation of *V. parahaemolyticus* clinical and environmental strains. *V. parahaemolyticus* strains have been shown to exhibit natural phenotypic variability such as phase variation, which is the transition from opaque (OP) to translucent (TR) colony morphologies (36). Following deletion of *mutS* in strain 17802 there was a significant change in colony phenotype from OP colony morphologies of the wild-type to all TR colonies for the Δ*mutS* strain (Table 4.7). Among the 64 clinical and 78 environmental strains examined for increased spontaneous mutation, 18 environmental strains (23%) and one clinical strain (1.2%) had a significant number of TR colonies compared to OP colonies (≥10% TR CFU/ml) (Table 4.7). Among the environmental strains exhibiting phase variation were all of the natural mutator strains with percentages of TR CFU/ml ranging from 28-100% (Table 4.7). The additional 14 environmental strains exhibiting phase variation had between 10-88% TR CFU/ml with an average of 47% TR CFU/ml (Table 4.7). The colonies of the Δ*mutS* strain had visibly less opacity than other TR colonies (data not shown). Four of the environmental strains exhibited three distinct colony morphologies, OP and TR, and a more translucent version of TR.

Table 4.7. Percentage of translucent (TR) colonies of *V. parahaemolyticus* clinical and environmental strains

Strain[a]	% TR CFU/ml (SD)[bc]
clinical	
RIMD	0.0 (0.0)
17802	0.5 (0.0)
17802 *ΔmutS*	**99.5 (0.0)[b]**
K1223	0.0 (0.0)
F7979	18.3 (0.1)
environmental	
22702	49.1 (0.1)
J-C1-5	**100 (0.0)**
J-C1-39	**28.9 (0.2)**
J-C2-27	**74.6 (0.4)**
J-C2-29	0.0 (0.0)
J-C2-34	**28.9 (0.1)**
J-C2-28	71.4 (0.4)[c]
J-C2-8a	88.3 (0.1)[c]
S-M2-2-B3	30.7 (0.1)
S-M2-3-B3	76.1 (0.0)[c]
S-M2-10-B3	87.3 (0.0)
AF110	66.7 (0.3)[c]
SG101	14.2 (0.1)
SG111	48.5 (0.1)
SG161	10.1 (0.1)
SG176	33.1 (0.2)
SG285	16.4 (0.2)
SG364	25.9 (0.0)
SG358	**26.4 (0.1)**

[a]Mutator strains are indicated in bold.
[b]Values represent the percentage of CFU/ml that are translucent (TR) determined from five or more independent replicates with the standard deviation (SD) indicated in parentheses.
[c]Three colony morphologies were observed: opaque (OP), translucent (TR), and more translucent.

Discussion

The loss of DNA repair mechanisms such as MMR may increase rates of bacterial diversification and facilitate more rapid niche expansion (15, 26). In this study, we detected a greater occurrence of mutators among *V. parahaemolyticus* environmental isolates compared to clinical strains. The finding of a higher frequency of mutators among environmental strains is in surprising contrast to previous reports for other bacteria that showed a greater frequency of mutators among clinical strains (3, 29, 44). The frequency of mutators identified among the *Vibrio* environmental strains in this study was approximately 7%, which was higher than the 1% identified for non-pathogenic *E. coli* strains (3). Mutators are usually identified in pathogenic populations, representing 2.6% of *E. coli* and *S. enterica* pathogens (29) and 19.5% of *P. aeruginosa* pathogens (44), while typically only 1% of non-pathogenic strains. An exception are the *E. coli* commensal strains that had higher mutation frequencies than *E. coli* clinical strains (3). The greater frequency of mutators observed among clinical bacteria is often linked to use of antibiotics for treatment of chronic infections. For example, cystic fibrosis patients have been shown to be infected with not only *P. aeruginosa* mutators, but also mutators of *H. influenzae* (58). This suggests that the stress of certain pathogenic niches or the prolonged use of antibiotics to treat chronic infections may induce mutations and loss of MMR in diverse bacteria. A possible explanation for the lack of mutators identified among the *V. parahaemolyticus* clinical strains in this study may be the self-limiting nature of *V. parahaemolyticus* infections and the infrequent use of antibiotics for treatment of these infections (6). *Vibrios* isolated from oysters exhibit resistance to numerous antibiotics including rifampin, ampicillin, nalidixic acid, and ciprofloxacin

(19). The increase in resistance to Cip demonstrated for the Δ*mutS* strain indicates a role of inactivation of MMR for increased resistance to an antibiotic used to treat severe *Vibrio* infections (54). A previous study has reported that mutations in *gyrA* or *parC* of *V. parahaemolyticus* can confer low-level resistance to Cip (42), while mutations in both *gyrA* and *parC* resulted in higher levels of resistance (42). The increased resistance to Cip we report for the Δ*mutS* strain, one of the *V. parahaemolyticus* mutator strains, and the *V. campbellii* mutator strain may have resulted from the accumulation of mutations in both *gyrA* and *parC* while the remaining mutator strains may not have mutations in both genes. The occurrence of mutators in *V. parahaemolyticus* populations from environmental sources may be one possible explanation for the increased observation of antibiotic-resistant *Vibrios* in the environment such as that recently reported for isolates cultured from oysters (19). In addition, the stress of the environmental niche may play a role in the emergence of *Vibrio* natural mutator strains. The majority of the *Vibrio* mutators in this study were isolated from sediment of the rhizosphere of *Juncus* spp. occurring in the high intertidal zone (1), which may undergo greater fluctuations in salinity, temperature, and UV exposure. In contrast, the *V. campbellii* mutator was isolated from water associated with *Spartina* spp. from the middle to low intertidal zones. The greater frequency of mutators associated with *Juncus* spp. suggests that the level of environmental stress may increase the occurrence of mutator strains in certain environmental niches. Further studies are necessary to investigate the role of environmental stress for the emergence of *Vibrio* mutators.

A previous study has reported the horizontal transfer and recombination of MMR genes as evidence of the inactivation and reacquisition of functional MMR genes in *E.*

coli (8). *Oenococcus oeni* and *O. kitaharae* were shown to lack *mutS* and *mutL*, which may account for their more rapid evolution relative to other *mutSL*-containing *Lactobacillales* as determined by 16S phylogenetic analysis (32). Inactivation of MMR increased the frequency of transition mutations in *rpoB* from Rif^r colonies of *Oenococcus* spp. (32), *B. anthracis* (61), and *E. coli* (4). Our results also showed that inactivation of *V. parahaemolyticus* MMR increased the frequency of transition mutations in *rpoB*. In addition, we found evidence of a role of mutator phenotypes for increased nucleotide diversity in select genes (*mutS* and *rpoB*). In contrast, variability at *recA* and *rpoA* were similar in the mutator and wild-type strains. The increase of nucleotide diversity of select genes in mutator compared to wild-type *V. parahaemolyticus* strains could instead reflect the diversity of environmental compared to clinical strains that has been previously reported (33). However, the fact that diversities of *recA* and *rpoA* alleles of mutator and wild-type strains were similar to each other indicates that the increased nucleotide diversity in *mutS* and *rpoB* alleles could be directly attributed to the increase in spontaneous mutation or recombination associated with the mutator phenotype. Overall, the levels of selective constraint observed in *V. parahaemolyticus* variability (π_{nonsyn}/π_{syn}) and nucleotide substitution compared to V. harveyi (K_a/K_s) indicates that the genes examined here are all under strong selective constraint. The increase of nucleotide variability in *mutS* and *rpoB* thus probably reflects increases in underlying mutation rates rather than an increase of loss-of-function mutations. Although there were no frameshift mutations or nonsense mutations identified in the mutator *mutS* sequences, alleles from several strains had non-synonymous substitutions that could have altered the function of the MutS protein. Likewise, the absence of frameshifts and nonsense mutations in the

mutator *mutS* alleles was reported in a previous study that examined the inactivation of *mutS* and *mutL* of *S. aureus* (46). Several of the *S. aureus* mutator MMR genes had only a few non-synonymous substitutions while other alleles had no substitutions at all (46).

Further analysis of the housekeeping genes revealed there was strong evidence of recombination events in *recA, rpoB,* and *mutS*. MLST of *V. parahaemolyticus* showed that while the pandemic strains are clonal they may have originated from diverse strains that had undergone high levels of recombination (16). The MLST study also showed that many of the non-synonymous changes in the housekeeping genes were the result of recombination rather than mutation. Of the seven housekeeping genes examined in the MLST study, *recA* had alleles with variation resulting from both mutational and recombinational changes (16). We have previously shown that *recA* may undergo horizontal transfer mediated by mobile genetic elements such as plasmids (20). We identified a nearly complete *recA* allele on a plasmid isolated from an environmental *Vibrio* strain most closely related to *V. mediterranei* (20). Thus, MMR-deficient strains may have increased likelihood of recombination of horizontally-acquired alleles such as a plasmid-encoded *recA* from other *Vibrio spp.* Although we have demonstrated in this study that the spontaneous mutation of *V. parahaemolyticus* mutators does not increase the selective pressure acting on housekeeping genes, inactivation of MMR has been previously shown to increase the frequency of recombination events (39, 46).

Recent studies have shown the genetic diversity of environmental *Vibrios* may correlate with the niche they occupy (24). A comparative genome hybridization of *V. cholerae* environmental strains revealed there were distinct populations that varied relative to changing environmental parameters such as nutrient composition, salinity, and

temperature (25). Mechanisms influencing the diversification of *V. parahaemolyticus* environmental strains may depend on abiotic and biotic factors of their environmental niche such as nutrient availability, oxidative stress, temperature fluctuations, competitive interactions, and the presence of mutator strains. For example, mutations could accumulate in genes such as *opaR*, which encodes a quorum sensing transcriptional regulator (36). OpaR is one of several proteins that controls the production of capsular polysaccharide that gives a colony an opaque appearance (17). *V. parahaemolyticus* strains with mutations in *opaR* exhibit delayed attachment and produce biofilms that have more distinct microcolonies and channels than observed in wild-type biofilms (11). OpaR mutants have been shown to naturally occur (36); however, the frequency of mutation of OpaR among *V. parahaemolyticus* clinical and environmental strains and the mechanisms and conditions contributing to the inactivation of OpaR have not been described. Inactivation of MMR has previously been shown to increase the frequency of mutation of the quorum sensing regulator, LasR, of *P. aeruginosa* (30) and we postulate a similar possibility for OpaR. In addition, *N. meningitidis* serogroup A isolates were shown to have increased mutation frequencies and phase variation (48, 49). *V. vulnificus* strains reported to have environmental genotypes were shown to have greater phase variation from OP to TR colony morphologies compared to strains with genotypes denoted as clinical (22). *V. vulnificus* phase variation was independent of quorum sensing and due instead to the disruption of the stationary phase regulator *rpoS* (22). Further investigation is underway to characterize the role of inactivation of MMR for accumulation of mutations in genes such as *rpoS* or *opaR* that could increase phenotypic diversity of *V. parahaemolyticus*.

In summary, we report that inactivation of the MMR gene *mutS* of *V. parahaemolyticus* increases the accumulation of spontaneous mutations conferring resistance to Rif and Cip. The higher frequency of transition mutations among mutator strains and increased nucleotide diversity in mutator *mutS* and *rpoB* alleles relative to wild-type alleles indicates a contribution of mutator phenotypes to the diversity of *V. parahaemolyticus*. Each of the housekeeping genes examined was under strong purifying selection; however, *mutS* and *rpoB* alleles of mutator strains had significantly more synonymous and nonsynonymous nucleotide changes compared to the wild-type alleles. Inactivation of MMR may also contribute to increased diversity of natural populations of *V. parahaemolyticus* by lowering barriers to recombination. The greater frequency of natural mutators among the *V. parahaemolyticus* environmental strains examined in this study may result from significant fluctuations in nutrients, salinity, and temperature these strains would experience in their environmental niche. Our study provides the first characterization of a higher frequency of mutators among *Vibrio* environmental strains and provides evidence that inactivation of MMR increases the diversity of *V. parahaemolyticus*. Research is ongoing to determine whether environmental stress increases the frequency of natural transformation and recombination for *V. parahaemolyticus* mutator strains.

Acknowledgements

I would like to thank D. Kennedy and S. Chen for laboratory assistance. Also, I would like to thank S. Yi for guidance on sequence analysis and contributions to the manuscript.

References

1. **Bagwell, C. E., Y. M. Piceno, A. Ashburne-Lucas, and C. R. Lovell.** 1998. Physiological diversity of the rhizosphere diazotroph assemblages of selected salt marsh grasses. Appl. Environ. Microbiol. **64:**4276-4282.
2. **Bhakat, K. K., S. Sharma, and J. Das.** 1999. The *mutK* gene of *Vibrio cholerae*: a new gene involved in DNA mismatch repair. J. Bacteriol. **181:**879-883.
3. **Bjedov, I., O. Tenaillon, B. Gérard, V. Souza, E. Denamur, M. Radman, F. Taddei, and I. Matic.** 2003. Stress-induced mutagenesis in bacteria. Science **300:**1404-1409.
4. **Choy, H. E., and R. G. Fowler.** 1985. The specificity of base-pair substitution induced by the MutL and MutS mutators in *Escherichia coli*. Mutat Res **142:**93-97.
5. **Criminger, J. D., T. H. Hazen, P. A. Sobecky, and C. R. Lovell.** 2007. Nitrogen fixation by *Vibrio parahaemolyticus* and its implications for a new ecological niche. Appl. Environ. Microbiol. **73:**5959-5961.
6. **Daniels, N. A., L. MacKinnon, R. Bishop, S. Altekruse, B. Ray, R. M. Hammond, S. Thompson, S. Wilson, N. H. Bean, P. M. Griffin, and L. Slutsker.** 2000. *Vibrio parahaemolyticus* infections in the United States, 1973-1998. J. Infect. Dis. **181:**1661-1666.
7. **Dehio, C., and M. Meyer.** 1997. Maintenance of broad-host-range incompatibility group P and group Q plasmids and transposition of Tn5 in *Bartonella henselae* following conjugal plasmid transfer from *Escherichia coli*. J Bacteriol **179:**538-540.
8. **Denamur, E., G. Lecointre, P. Darlu, O. Tenaillon, C. Acquaviva, C. Sayada, I. Sunjevaric, R. Rothstein, J. Elion, F. Taddei, M. Radman, and I. Matic.** 2000. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. Cell **103:**711-721.
9. **Denamur, E., and I. Matic.** 2006. Evolution of mutation rates in bacteria. Mol. Microbiol. **60:**820-827.
10. **DePaola, A., J. Ulaszek, C. A. Kaysner, B. J. Tenge, J. L. Nordstrom, J. Wells, N. Puhr, and S. M. Gendel.** 2003. Molecular, serological, and virulence characteristics of *Vibrio parahaemolyticus* isolated from environmental, food, and clinical sources in North America and Asia. Appl. Environ. Microbiol. **69:**3999-4005.
11. **Enos-Berlage, J. L., Z. T. Guvener, C. E. Keenan, and L. L. McCarter.** 2005. Genetic determinants of biofilm development of opaque and translucent *Vibrio parahaemolyticus*. Mol. Microbiol. **55:**1160-1182.
12. **Ferenci, T.** 2003. What is driving the acquisition of *mutS* and *rpoS* polymorphisms in *Escherichia coli*? Trends Microbiol. **11:**457-461.
13. **Friedhoff, P., B. Sheybani, E. Thomas, C. Merz, and A. Pingoud.** 2002. *Haemophilus influenzae* and *Vibrio cholerae* genes for *mutH* are able to fully complement a *mutH* defect in *Escherichia coli*. FEMS Microbiol. Lett. **208:**123-128.
14. **Fujino, T., T. Miwatani, J. Yasuda, M. Kondo, Y. Takeda, Y. Akita, K. Kotera, M. Okada, H. Nishimune, Y. Shimizu, T. Tamura, and Y. Tamura.**

1965. Taxonomic studies on the bacterial strains isolated from cases of "shirasu" food-poisoning (*Pasteurella parahaemolytica*) and related microorganisms. Biken J. **8:**63-71.

15. **Giraud, A., I. Matic, O. Tenaillon, A. Clara, M. Radman, M. Fons, and F. Taddei.** 2001. Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. Science **291:**2606-8.
16. **González-Escalona, N., J. Martinez-Urtaza, J. Romero, R. T. Espejo, L. Jaykus, and A. DePaola.** 2008. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. J. Bacteriol. **190:**2831-2840.
17. **Güvener, Z. T., and L. L. McCarter.** 2003. Multiple regulators control capsular polysaccharide production in *Vibrio parahaemolyticus*. J Bacteriol **185:**5431-41.
18. **Hall, T. A.** 1999. BioEdit: a user-friendly biological sequence alignment editor and anlysis program for Windows 95/98/NT. Nucleic Acids Symp Ser **41:**95-98.
19. **Han, F., R. D. Walker, M. E. Janes, W. Prinyawiwatkul, and B. Ge.** 2007. Antimicrobial susceptibility of *Vibrio parahaemolyticus* and *Vibrio vulnificus* from Louisiana gulf and retail raw oysters. Appl. Environ. Microbiol. **73:**7096-7098.
20. **Hazen, T. H., D. Wu, J. A. Eisen, and P. A. Sobecky.** 2007. Sequence characterization and comparative analysis of three plasmids isolated from environmental *Vibrio* spp. Appl. Environ. Microbiol. **73:**7703-7710.
21. **Herbelin, C. J., S. C. Chirillo, K. A. Melnick, and T. S. Whittam.** 2000. Gene conservation and loss in the *mutS-rpoS* genomic region of pathogenic *Escherichia coli*. J. Bacteriol. **182:**5381-5390.
22. **Hilton, T., T. Rosche, B. Froelich, B. Smith, and J. Oliver.** 2006. Capsular polysaccharide phase variation in *Vibrio vulnificus*. Appl. Environ. Microbiol. **72:**6986-6993.
23. **Hudson, R. R., and N. L. Kaplan.** 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:**147-64.
24. **Hunt, D. E., L. A. David, D. Gevers, S. P. Preheim, E. J. Alm, and M. F. Polz.** 2008. Resource partitioning and sympatric differentiation among closely related bacterioplankton. Science **320:**1081-1085.
25. **Keymer, D. P., M. C. Miller, G. K. Schoolnik, and A. B. Boehm.** 2007. Genomic and phenotypic diversity of coastal *Vibrio cholerae* strains is linked to environmental factors. Appl. Environ. Microbiol. **73:**3705-3714.
26. **Kivisaar, M.** 2003. Stationary phase mutagenesis: mechanisms that accelerate adaptation of microbial populations under environmental stress. Environ. Microbiol. **5:**814-827.
27. **Kovach, M. E., E. W. Phillips, P. H. Elzer, R. M. Roop 2nd, and K. M. Peterson.** 1994. pBBR1MCS: a broad-host-range cloning vector. Biotechniques **16:**800-802.
28. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinf. **5:**150-163.

29. **LeClerc, J. E., L. Baoguang, W. L. Payne, and T. A. Cebula.** 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. Science **274:**1208-1211.
30. **Lújan, A. M., A. J. Moyano, I. Segura, C. E. Argaraña, and A. M. Smania.** 2007. Quorum-sensing-deficient (*lasR*) mutants emerge at high frequency from a *Pseudomonas aeruginosa mutS* strain. Microbiology **153:**225-237.
31. **Makino, K., K. Oshima, K. Kurokawa, K. Yokoyama, T. Uda, K. Tagomori, Y. Iijima, M. Najima, M. Nakano, A. Yamashita, Y. Kubota, S. Kimura, T. Yasunaga, T. Honda, H. Shinagawa, M. Hattori, and T. Iida.** 2003. Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*. Lancet **361:**743-749.
32. **Marcobal, A. M., D. A. Sela, Y. I. Wolf, K. S. Makarova, and D. A. Mills.** 2008. The role of hypermutability in the evolution of the genus *Oenococcus*. J. Bacteriol. **190:**564-570.
33. **Martinez-Urtaza, J., A. Lozano-Leon, A. DePaola, M. Ishibashi, K. Shimada, M. Nishibuchi, and E. Liebana.** 2004. Characterization of pathogenic *Vibrio parahaemolyticus* isolates from clinical sources in Spain and comparison with Asian and North American pandemic isolates. J. Clin. Microbiol. **42:**4672-8.
34. **Martinez-Urtaza, J., A. Lozano-Leon, A. Viña-Feas, J. de Novoa, and O. Garcia-Martin.** 2006. Differences in the API 20E biochemical patterns of clinical and environmental *Vibrio parahaemolyticus* isolates. FEMS Microbiol. Lett. **255:**75-81.
35. **Matic, I., C. Rayssiguier, and M. Radman.** 1995. Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. Cell **80:**507-515.
36. **McCarter, L. L.** 1998. OpaR, a homolog of *Vibrio harveyi* LuxR, controls opacity of *Vibrio parahaemolyticus*. J. Bacteriol. **180:**3166-3173.
37. **McLaughlin, J. B., A. DePaola, C. A. Bopp, K. A. Martinek, N. P. Napolilli, C. G. Allison, S. L. Murray, E. C. Thompson, M. M. Bird, and J. P. Middaugh.** 2005. Outbreak of *Vibrio parahaemolyticus* gastroenteritis associated with Alaskan oysters. N. Engl. J. Med. **353:**1463-70.
38. **Meador, C. E., M. A. Parsons, C. A. Bopp, P. Gerner-Smidt, J. A. Painter, and G. J. Vora.** 2007. Virulence gene- and pandemic group-specific marker profiling of clinical *Vibrio parahaemolyticus* isolates. J. Clin. Microbiol. **45:**1133-1139.
39. **Meier, P., and W. Wackernagel.** 2005. Impact of *mutS* inactivation on foreign DNA acquisition by natural transformation in *Pseudomonas stutzeri*. J. Bacteriol. **187:**143-154.
40. **Milton, D. L., R. O'Toole, P. Hörstedt, and H. Wolf-Watz.** 1996. Flagellin A is essential for the virulence of *Vibrio anguillarum*. J. Bacteriol. **178:**1310-1319.
41. **Nei, M.** 1987. Molecular Evolutionary Genetics, vol. Columbia University Press, New York.
42. **Okuda, J., E. Hayakawa, M. Nishibuchi, and T. Nishino.** 1999. Sequence analysis of the *gyrA* and *parC* homologues of a wild-type strain of *Vibrio parahaemolyticus* and its fluoroquinolone-resistant mutants. Antimicrob. Agents Chemother. **43:**1156-1162.

43. **Oliver, A., F. Baquero, and J. Blazquez.** 2002. The mismatch repair system (*mutS, mutL,* and *uvrD* genes) in *Pseudomonas aeruginosa*: molecular characterization of naturally occurring mutants. Mol. Microbiol. **43:**1641-1650.
44. **Oliver, A., R. Cantón, P. Campo, F. Baquero, and J. Blázquez.** 2000. High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. Science **288:**1251-1253.
45. **Parsons, M. B., K. L. Cooper, K. A. Kubota, N. Puhr, S. Simington, P. S. Calimlim, D. Schoonmaker-Bopp, C. Bopp, B. Swaminathan, P. Gerner-Smidt, and E. M. Ribot.** 2007. PulseNet USA standardized pulsed-field gel electrophoresis protocol for subtyping of *Vibrio parahaemolyticus*. Foodborne Pathog. Dis. **4:**285-92.
46. **Prunier, A. L., and R. Leclercq.** 2005. Role of *mutS* and *mutL* genes in hypermutability and recombination in *Staphylococcus aureus*. J. Bacteriol. **187:**3455-3464.
47. **Rayssiguier, C., Thaler, D. S., and Radman, M.** 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature **342:**396-401.
48. **Richardson, A. R., and I. Stojiljkovic.** 2001. Mismatch repair and the regulation of phase variation in *Neisseria meningitidis*. Mol Microbiol **40:**645-655.
49. **Richardson, A. R., Z. Yu, T. Popovic, and I. Stojiljkovic.** 2002. Mutator clones of *Neisseria meningitidis* in epidemic serogroup A disease. Proc Nat Acad of Sci USA **99:**6103-6107.
50. **Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas.** 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:**2496-7.
51. **Sawyer, S.** 1989. Statistical tests for detecting gene conversion. Mol. Biol. Evol. **6:**526-538.
52. **Scally, M., E. L. Schuenzel, R. Stouthamer, and L. Nunney.** 2005. Multilocus sequence type system for the plant pathogen *Xylella fastidiosa* and relative contributions of recombination and point mutation to clonal diversity. Appl. Environ. Microbiol. **71:**8491-8499.
53. **Schofield, M. J., and P. Hsieh.** 2003. DNA mismatch repair: molecular mechanisms and biological function. Annu Rev Microbiol **57:**579-608.
54. **Tang, H., M. Chang, W. Ko, K. Huang, C. Lee, and Y. Chuang.** 2002. In vitro and in vivo activities of newer fluoroquinolines against *Vibrio vulnificus*. Antimicrob. Agents Chemother. **46:**3580-3584.
55. **Tarr, C. L., J. S. Patel, N. D. Puhr, E. G. Sowers, C. A. Bopp, and N. A. Strockbine.** 2007. Identification of *Vibrio* isolates by a multiplex PCR assay and *rpoB* sequence determination. J. Clin. Microbiol. **45:**134-40.
56. **Thompson, F. L., T. Iida, and J. Swings.** 2004. Biodiversity of *Vibrios*. Microbiol. Mol. Biol. Rev. **68:**403-431.
57. **Thompson, J. R., M. A. Randa, L. A. Marcelino, A. Tomita-Mitchell, E. Lim, and M. F. Polz.** 2004. Diversity and dynamics of a north Atlantic coastal *Vibrio* community. Appl. Environ. Microbiol. **70:**4103-4110.

58. **Watson, M. E., J. L. Burns, and A. L. Smith.** 2004. Hypermutable *Haemophilus influenzae* with mutations in *mutS* are found in cystic fibrosis sputum. Microbiology-Sgm **150:**2947-2958.
59. **Watterson, G. A.** 1975. On the number of segregation sites. Theoret. Popul. Biol. **7:**256-276.
60. **Zahrt, T. C., N. Buchmeier, and S. Maloy.** 1999. Effect of *mutS* and *recD* mutations on *Salmonella* virulence. Infect. Immun. **67:**6168-6172.
61. **Zeibell, K., S. Aguila, V. Y. Shi, A. Chan, H. Yang, and J. H. Miller.** 2007. Mutagenesis and repair in *Bacillus anthracis*: the effect of mutators. J. Bacteriol. **189:**2331-2338.

CHAPTER 5

CONCLUSIONS

The research in chapters 2-4 focused on examining the contribution of horizontal gene transfer to the diversity of the pathogenic marine bacterium *V. parahaemolyticus.* The main research objective was to determine how genetic elements that may be horizontally transferred, and molecular mechanisms that lead to increased mutation and recombination contribute to the differences in genetic diversity identified among *V. parahaemolyticus* clinical and environmental strains. In chapter 2, I compared the diversity of *V. parahaemolyticus* strains present in the environment to strains that were isolated from disease-associated samples. The research in chapter 3 focused on examining the diversity of genetic elements associated with *V. parahaemolyticus* clinical and environmental strains for evidence of the contribution of MGEs to strain diversity. Finally, in chapter 4 I demonstrated that molecular mechanisms such as DNA repair processes can be naturally altered to result in increases in the contribution of mutation and recombination to strain diversity.

Summary of Dissertation Research Findings

Diversity of *V. parahaemolyticus* clinical and environmental strains

Research in chapter 2 characterized the diversity of *V. parahaemolyticus* clinical strains isolated from disease-associated samples to strains isolated from environmental samples. The diversity of *V. parahaemolyticus* environmental strains isolated from sediment, water, and oysters was examined for co-occurring as well as geographically

isolated strains. There were strains that were most related to geographically co-occurring strains and separate strains that showed relatedness to strains from each sample location indicating a widely-distributed strain type. Comparison of the relatedness of the *V. parahaemolyticus* environmental strains to the clinical strains isolated from domestic disease events indicated that the disease-causing strains and environmental strains formed two mostly unique lineages. An exception to the separation of the *V. parahaemolyticus* clinical and environmental strains were 10 environmental strains that grouped with clinical strains and seven clinical strains that grouped with environmental strains.

In addition, in chapter 2 I demonstrated the use of whole-cell MALDI-TOF MS analysis for identification of *V. parahaemolyticus* clinical and environmental strains. Whole-cell MALDI analysis allowed us to distinguish *V. parahaemolyticus* strains from related *Vibrios* including *V. harveyi* and *V. campbellii.* The whole-cell MALDI technique described allowed rapid and reliable identification of *V. parahaemolyticus* that is faster than other frequently used biochemical and molecular techniques that are currently employed to identify *V. parahaemolyticus.* The unique MALDI fingerprints of the *V. parahaemolyticus* O3:K6 and O4:K12 clonal strains indicated that MALDI may possibly be used to examine the evolution of disease-causing strains that are closely related and thus difficult to distinguish based on other molecular techniques.

Diversity of *Vibrio* plasmids and their role in the evolution of *V. parahaemolyticus* clinical and environmental strains

In chapter 3, I described the sequences of *Vibrio* plasmids isolated from diverse environmental *Vibrios*. These plasmids were isolated from three diverse environmental *Vibrios* and there was no sequence conservation among the plasmids indicating that

plasmids associated with environmental *Vibrios* have significant diversity. This was the first characterization of an extrachromosomal element isolated from each of the three environmental *Vibrios* examined. The plasmids encoded phage-like genes and one of the plasmids encoded genes of the enterobacterial phage P1 indicating horizontal gene transfer occurred between a *Vibrio* plasmid and a phage that is associated with bacteria present in the gut. In addition, there were housekeeping genes identified on one of the plasmids analyzed indicating that plasmids contribute to the evolution of housekeeping gene sequences among *Vibrios*.

In addition, I described an environmental plasmid family that was detected among *V. parahaemolyticus, V. harveyi,* and *V. campbellii* strains isolated from environmental samples. Southern hybridization of a *rep* probe generated from a sequenced plasmid to other medium to large plasmids isolated from *V. parahaemolyticus, V. campbellii,* and *V. harveyi* strains revealed a conserved rep-type that was widely-distributed among related *Vibrios* from geographically co-occurring and separated environmental samples. In addition, there was a large genetic element (~86-kb) detected among three *V. parahaemolyticus* clinical strains and four *V. parahaemolyticus* environmental strains isolated from GA and NC. Structural analysis of these putative plasmids by restriction endonuclease digestion indicated one element from a clinical strain and of the two elements from environmental strains had identical restriction profiles. Sequence characterization of a 28.8-kb *V. parahaemolyticus* plasmid revealed this genetic element possessed a putative bacteriocin-encoding gene that has previously been identified on the genomic island VPaI-6 that has been detected only among *V. parahaemolyticus* pandemic strains (6).

Molecular mechanisms of the evolution of *V. parahaemolyticus*

The research in chapter 4 illustrated the contribution of the inactivation of DNA repair mechanisms to the increased accumulation of mutation and increased frequency of recombination to the evolution of *V. parahaemolyticus* strain diversity. I demonstrated the occurrence of *V. parahaemolyticus* natural mutator strains among environmental populations and showed there were no natural mutators detected among the *V. parahaemolyticus* clinical strains examined. As demonstrated for other bacteria examined to date, inactivation of the MMR gene *mutS* lead to an increase in the accumulation of spontaneous mutations. Furthermore, there was evidence of a contribution of inactivation of MMR to increased mutation in certain select genes of mutator strains examined, as illustrated by a higher frequency of transition mutations.

Broader Implications of Dissertation Research

Environmental reservoirs of *V. parahaemolyticus* as an emerging pathogen

The research findings of chapter 2 demonstrated that some *V. parahaemolyticus* strains in the environment are genetically related to disease-causing strains. Although the environmental strains lacked the main virulence-associated genes, they may possess other genes required for host colonization and establishment of infection and upon acquisition of virulence genes they may emerge as pathogens under permissive conditions. These findings are the first to show an evolutionary relationship between *V. parahaemolyticus* clinical and environmental strains with the identification of intermediate pathogens and emerging clonal pandemic strains. Previous studies that examined the diversity of *V. parahaemolyticus* clinical and environmental strains reported significant genetic diversity

among environmental strains compared to the closely related clinical strains. The research findings in chapter 2 advance the understanding of the relationships among *V. parahaemolyticus* clinical and environmental strains. The identification of environmental strains that exhibit relatedness to clinical strains demonstrates that sediment and water in coastal environments could serve as a possible reservoir for emerging pathogens that upon acquisition of the appropriate virulence-associated genes would pose the potential for initiating a disease outbreak. This information indicates the necessity to study environmental populations and to include these populations of strains when developing models for disease outbreak monitoring and prediction even though the strains may not possess the known virulence-associated genes.

In addition, I demonstrated in chapter 2 that whole-cell MALDI-TOF MS analysis is a powerful tool for the rapid identification and classification of *V. parahaemolyticus* strains. Application of whole-cell MALDI-TOF MS analysis would significantly reduce the amount of time required to identify the *Vibrio* spp. responsible for ongoing disease outbreaks. This research is the first demonstration of the use of whole-cell MALDI-TOF MS analysis for the identification of *Vibrios*. The whole-cell MALDI-TOF MS analysis is highly reproducible for the identification of *V. parahaemolyticus.* Furthermore, MALDI-TOF MS is a more rapid and relatively easy method for the identification of *V. parahaemolyticus* than other currently used methods such as MLST and PFGE. Whole-cell MALDI-TOF MS analysis would be a valuable tool for the preliminary detection of *V. parahaemolyticus* in a sample and could be used in lieu of biochemical tests to identify strains isolated to pure culture that are presumed to be *V. parahaemolyticus.* MALDI-TOF MS results could then be compared with results from PFGE analysis in order to

characterize strain diversity. In addition, whole-cell MALDI-TOF MS analysis provided unique fingerprints for some of the *V. parahaemolyticus* disease-causing strains analyzed that are clonal based on PFGE and MLST analysis. The unique fingerprints of the clonal strains that are provided by whole-cell MALDI-TOF MS analysis could be further investigated for the development of biomarker peaks used to identify the origin of the pandemic clones. Furthermore, whole-cell MALDI-TOF MS analysis would be a valuable method to monitor the evolution and spread of the clonal pandemic strains.

Contribution of *Vibrio* MGEs to the evolution of disease-causing and non-pathogenic environmental strains

The identification of a significant number of P1 phage genes on a *Vibrio* plasmid provided evidence that phage and plasmids facilitate inter-species gene transfer between potentially pathogenic *Vibrios* and co-occurring enteric bacteria. In addition, plasmid analysis revealed that the plasmid-mediated horizontal transfer of housekeeping genes such as *recA* confounds the use of *recA* as a conserved phylogenetic marker for identification of *Vibrios* as previously performed (3, 4). Furthermore, this research provided the first glimpse into the genetic content of plasmids isolated from diverse *Vibrio* environmental strains.

In addition, the research in chapter 3 demonstrated the occurrence of related plasmids among *V. parahaemolyticus* and the closely related *Vibrios V. harveyi* and *V. campbellii*. Structurally conserved genetic elements were identified from one *V. parahaemolyticus* clinical strain and two *V. parahaemolyticus* environmental strains. This finding indicates that *V. parahaemolyticus* genetic elements may be involved in gene exchange among clinical and environmental strains or that the clonal O4:K12 strains

that are frequently associated with infections in the United States emerged from *V. parahaemolyticus* environmental populations in coastal regions of the U.S. following the acquisition of virulence-associated genes. In addition, research in chapter 3 demonstrates the contribution of *V. parahaemolyticus* plasmids to gene exchange between clonal pandemic strains and environmental strains. The identification of genomic island gene typically associated with *V. parahaemolyticus* O3:K6 clonal pandemic strains on a plasmid isolated from a *V. parahaemolyticus* environmental strain showed that plasmids may facilitate the exchange of genes between pandemic strains and strains in the environment. It is not clear whether the plasmid is transferring the genomic island gene from a pandemic strain to an environmental strain or whether plasmids may have been involved in the formation of the VPaI-6 genomic island. As demonstrated in chapter 2, the plasmid-bearing host strain *V. parahaemolyticus* 22702 forms a group within the clinical strain lineage of the *V. parahaemolyticus* phylogeny suggesting that this strain may be an emerging pathogen. This research is the first description of the contribution of plasmids to horizontal transfer of genes associated with the *V. parahaemolyticus* pandemic strains.

Evolution of *V. parahaemolyticus* populations

Research described in chapter 2 reported the presence of *V. parahaemolyticus* disease-causing strains emerging from environmental strain reservoirs. The research of chapter 3 described the contribution of plasmids to the evolution of *Vibrio* housekeeping genes and the emergence of *V. parahaemolyticus* pandemic strains from environmental strains. In addition, the research findings in chapter 4 provide new insight into the contribution of molecular mechanisms such as DNA repair pathways to the diversity of

V. parahaemolyticus clinical and environmental strains. In chapter 4, I demonstrated that mutator strains that have high spontaneous mutation frequencies are present among *V. parahaemolyticus* environmental strains rather than clinical strains. There has been only one study that has previously identified natural mutators among environmental populations (2), highlighting the importance for further research into the role of mutator phenotypes for evolution during environmental stress. Inactivation of MMR not only increased the accumulation of transition mutations in select genes, but also increased the occurrence of antibiotic resistance. The natural inactivation of MMR among *V. parahaemolyticus* environmental strains may explain the recent detection of environmental strains that exhibit resistance to multiple antibiotics (1). In addition, the research presented in chapter 4 is the first identification of natural mutator strains among marine bacteria.

Future Research

Genome sequence analysis of *V. parahaemolyticus* strains

In chapter 2, I discussed the evolutionary relationships of numerous strains that would be excellent candidates for genome sequencing. In particular, genome sequence analysis of the emerging pandemic strain *V. parahaemolyticus* AF91 that I had isolated from an environmental sample from Florida would provide insight into environmental strains that may be in an intermediary stage prior to becoming pathogenic. Sequence analysis of the genome of AF91 would provide information on the genetic potential of the possible *V. parahaemolyticus* environmental reservoir.

In addition, several *V. parahaemolyticus* clinical strains that were isolated by the CDC from disease-associated samples exhibited unique virulence-associated gene profiles. The *V. parahaemolyticus* O4:K12 strains are frequently associated with outbreaks in the United States and are a clonal group believed to have emerged from the O3:K6 clonal pandemic strains (11). The *V. parahaemolyticus* strain K1461 is a representative O4:K12 strain that is clonally-related to other O4:K12 strains based on PFGE and MLST analysis. In addition, this strain possesses both thermostable direct hemolysins, *tdh* and *trh,* which is unlike the O3:K6 clonal pandemic strains that have only *tdh*. Also, I showed in chapter 3 that K1461 possessed an extrachromosomal element that was structurally related to two other elements of an identical size that were isolated from geographically separated environmental strains. Sequence analysis of the genome of this strain would provide further insight into the virulence-associated genes required for the pathogenic mechanism of *V. parahaemolyticus* strains.

Development of a whole-cell MALDI-TOF MS database for identification of *Vibrios*

The use of whole-cell MALDI-TOF MS analysis for identification of *V. parahaemolyticus* clinical and environmental strains could be further developed to include of a database of *Vibrio* MALDI-TOF MS fingerprints. In addition, the whole-cell MALDI-TOF MS method described in chapter 2 could be tested on environmental samples and disease-associated samples for the detection of multiple *Vibrio* spp. in a single sample. Theoretically, a mixed *Vibrio* spp. biological sample that also contained other prokaryotes and eukaryotic cells could be inoculated to a specific medium and cultured overnight then analyzed by whole-cell MALDI-TOF MS. Use of whole-cell MALDI-TOF MS analysis in this manner would significantly reduce the time required to

determine whether an environmental or disease-associated sample contained *V. parahaemolyticus.* Samples that test positive for *V. parahaemolyticus* could then be plated on selective media and subjected to biochemical tests to isolate a *V. parahaemolyticus* strain.

Further development of the whole-cell MALDI-TOF MS analysis method for clinical diagnostics of *V. parahaemolyticus* disease-associated strains could be used to distinguish pathogenic from non-pathogenic strains. Analysis of many more *V. parahaemolyticus* clinical and environmental strains by whole-cell MALDI-TOF MS could result in the identification of biomarker peaks that distinguish strains that are potential pathogens from those that are benign environmental strains that may be co-occurring with a disease-causing strain in a sample. This approach could be particularly useful for determining what strains actually possess incompletely characterized pathogenic mechanisms since recent studies have identified strains that lack previously identified virulence-associated genes (9).

Plasmid sequence analysis

Further investigation of the role of MGEs for the evolution of *V. parahaemolyticus* clinical and environmental strains is being examined by the characterization of two plasmids identified in chapter 3. Sequencing and characterization of the structurally conserved 86-kb extrachromosomal elements pK1461 of the *V. parahaemolyticus* O4:K12 clonal strains K1461 and pJ-C2-34 of the *V. parahaemolyticus* environmental strain is underway. Furthermore, I demonstrated in chapter 2 that the *V. parahaemolyticus* clinical strain that possessed the conserved genetic element is a clonal O4:K12 strain that has both thermostable direct hemolysins (*tdh* and *trh*). This strain

lacks the T3SS2α genes that are associated with the *tdh*$^+$ pandemic strains (13). A second T3SS2 system, designated T3SS2β, was recently identified for the *trh*$^+$/*tdh*$^-$ clinical strains (12). Although this conserved genetic element is not present in other O4:K12 strains analyzed, further characterization of this element will provide insight into the role of genetic elements in gene exchange among O4:K12 strains frequently involved in *V. parahaemolyticus* infections in the United States and environmental strains. Analysis of these elements would indicate whether they are plasmids, genomic islands, or phage. In addition, we would be able to analyze the protein-encoding genes and determine whether they might be involved in pathogenicity or environmental adaptation phenotypes. Differences in the G+C content of these plasmids compared to the 45% G+C content of the *V. parahaemolyticus* chromosomes might indicate whether these elements were recently horizontally acquired from another bacterium. Genetically conserved regions shared between the two plasmids would be analyzed by multiple sequence alignment with homologs identified by BLAST analysis to determine the identity of the plasmid-encoded sequences. In addition, we will examine the sequence conservation of the two sequenced plasmids to an additional structurally conserved plasmid isolated from *V. parahaemolyticus* environmental strain SG176 from GA. Comparison of the sequence identity of these plasmids would provide insight into whether they encode proteins that facilitate adaptation to the environment or confer pathogenicity. Analysis of these plasmids will also provide information regarding the evolution of *V. parahaemolyticus* MGEs and their role in HGT among non-pathogenic and disease-causing strains. A recent study revealed the contribution of plasmid-encoded genes from a 68-kb plasmid to the pathogenic mechanism of *V. vulnificus* Biotype II

strains that cause disease in eels (7). Analysis of the distribution of the virulence plasmid revealed that it was present in all Biotype II strains and not detected in Biotype I or III *V. vulnificus* strains (14).

Mismatch repair and strain evolution

Further studies to investigate the contribution of inactivation of MMR for diversification of *V. parahaemolyticus* would examine whether MMR deficient strains have increased recombination frequencies of divergent alleles as reported for other bacteria (10). In addition, further studies are needed to determine the relative contribution of mutator strains to the overall diversity of a population in the environment by increased frequencies of recombination and accumulation of mutations. Further studies should examine the frequency of natural mutators among other opportunistic marine pathogens and non-pathogenic bacteria, to understand the impact of environmental stress of the emergence of mutator strains in marine environments and the contribution of mutators to environmental reservoirs of antibiotic resistance. In addition to the increased antibiotic resistance we demonstrated for several of the *V. parahaemolyticus* mutator strains examined, other studies have shown multi-drug resistance of bacteria involved in chronic infections such as *Pseudomonas aeruginosa* (8) results from inactivation of MMR. A recent study demonstrated that approximately 24% of *V. parahaemolyticus* strains isolated from the environment are resistant to antibiotics (1) indicating a need to study the mechanisms and environmental conditions that lead to antibiotic resistance in natural populations of *V. parahaemolyticus.* In addition, further studies are needed to determine the mechanisms and the frequency of horizontal transfer

of antibiotic resistance determinants from *V. parahaemolyticus* to other opportunistic pathogens in the environment or a host.

References

1. **Baker-Austin, C., J. V. McArthur, R. C. Tuckfield, M. Najarro, A. H. Lindell, J. Gooch, and R. Stepanauskas.** 2008. Antibiotic resistance in the shellfish pathogen *Vibrio parahaemolyticus* isolated from the coastal water and sediment of Georgia and South Carolina, USA. J. Food Prot. **71:**2552-2558.
2. **Bjedov, I., O. Tenaillon, B. Gérard, V. Souza, E. Denamur, M. Radman, F. Taddei, and I. Matic.** 2003. Stress-induced mutagenesis in bacteria. Science **300:**1404-1409.
3. **Chowdhury, N. R., O. C. Stine, J. Glenn Morris, and G. B. Nair.** 2004. Assessment of evolution of pandemic *Vibrio parahaemolyticus* by multilocus sequence typing. J. Bacteriol. **42:**1280-1282.
4. **González-Escalona, N., J. Martinez-Urtaza, J. Romero, R. T. Espejo, L. Jaykus, and A. DePaola.** 2008. Determination of molecular phylogenetics of *Vibrio parahaemolyticus* strains by multilocus sequence typing. J. Bacteriol. **190:**2831-2840.
5. **Hilton, T., T. Rosche, B. Froelich, B. Smith, and J. Oliver.** 2006. Capsular polysaccharide phase variation in *Vibrio vulnificus*. Appl. Environ. Microbiol. **72:**6986-6993.
6. **Hurley, C. C., A. M. Quirke, F. J. Reen, and E. F. Boyd.** 2006. Four genomic islands that mark post-1995 pandemic *Vibrio parahaemolyticus* isolates. BMC Genom. **7:**104.
7. **Lee, C. T., C. Amaro, K. M. Wu, E. Valiente, Y. F. Change, S. F. Tsai, C. H. Chang, and L. I. Hor.** 2008. A common virulence plasmid in biotype 2 *Vibrio vulnificus* and its dissemination aided by a conjugal plasmid. J. Bacteriol. **190:**1638-48.
8. **Macia, M. D., D. Blanquer, B. Togores, J. Sauleda, J. L. Perez, and A. Oliver.** 2005. Hypermutation is a key factor in development of multiple-antimicrobial resistance in Pseudomonas aeruginosa strains causing chronic lung infections. Antimicrobial Agents and Chemotherapy **49:**3382-3386.
9. **Meador, C. E., M. A. Parsons, C. A. Bopp, P. Gerner-Smidt, J. A. Painter, and G. J. Vora.** 2007. Virulence gene- and pandemic group-specific marker profiling of clinical *Vibrio parahaemolyticus* isolates. J. Clin. Microbiol. **45:**1133-1139.
10. **Meier, P., and W. Wackernagel.** 2005. Impact of *mutS* inactivation on foreign DNA acquisition by natural transformation in *Pseudomonas stutzeri*. J. Bacteriol. **187:**143-154.
11. **Nair, G. B., T. Ramamurthy, S. K. Bhattacharya, B. Dutta, Y. Takeda, and D. A. Sack.** 2007. Global dissemination of *Vibrio parahaemolyticus* serotype O3:K6 and its serovariants. Clin. Microbiol. Rev. **20:**39-48.

12. **Okada, N., T. Iida, K. Park, N. Goto, T. Yasunaga, H. Hiyoshi, S. Matsuda, T. Kodama, and T. Honda.** 2009. Identification and characterization of a novel type III secretion system in *trh*-positive *Vibrio parahaemolyticus* strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level. Infect. Immun. **77:**904-913.
13. **Park, K. S., T. Ono, M. Rokuda, M. H. Jang, K. Okada, T. Iida, and T. Honda.** 2004. Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. Infect. Immun. **72:**6659-6665.
14. **Roid, F. J., and C. Amaro.** 2009. Plasmid diversity in *Vibrio vulnificus* biotypes. Microbiol.-SGM **155:**489-497.

VITA

TRACY HEATHER HAZEN

Hazen was born in San Juan, Puerto Rico where she lived until 5 years old when she moved with her family to Augusta, Georgia. Hazen is the daughter of Dr. Terry Hazen and Gayle Hazen and has a brother, Brooks Hazen. The Hazen family moved from Georgia to Northern California in 1998. Hazen graduated from Armijo High School in 2000 and attended the University of California, Davis earning a B.S. in Biological Sciences with an emphasis in Marine Biology in 2004. Hazen began studying for a Ph.D. in Biology at the Georgia Institute of Technology in the Fall of 2004. When not studying microbiology, Hazen enjoys photography, outdoor activities, and traveling.